RTS-0239

BB
M
J

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 March 2002 (21.03.2002)

PCT

(10) International Publication Number
WO 02/22817 A2

(51) International Patent Classification[7]: C12N 15/12, C07K 14/705, C12N 5/10, G01N 33/68, C07K 16/28, C12Q 1/68, C12N 15/11, A61K 48/00, 39/395, 31/7088

(21) International Application Number: PCT/EP01/10323

(22) International Filing Date: 7 September 2001 (07.09.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
00120370.2 16 September 2000 (16.09.2000) EP
01111658.9 14 May 2001 (14.05.2001) EP

(71) Applicant (*for all designated States except US*): **LION BIOSCIENCE AG** [DE/DE]; Im Neuenheimer Feld 515, 69120 Heidelberg (DE).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): **CASARI, Georg** ~~[AT/DE]; Hauptstrasse 313 (5), 69124 Eppelheim (DE).~~ **HÖFER, Michael** [DE/DE]; Im Weiher 61, 69121 Heidelberg (DE). **JACKSON, David** [IE/DE]; Schröderstrasse 61, 69120 Heidelberg (DE). **KRANZ, Harald** [DE/DE]; Schlesierstrasse 12, 69221 Dossenheim (DE). **OTTE, Kerstin** [DE/DE]; Biethstrasse 20, 69121 Heidelberg (DE). **REMMEL, Bettina** [DE/DE]; Birkenweg 8A, 69221 Dossenheim (DE). **SUCKOW, Jörg** [DE/DE]; Uhlandstrasse 5, 69221 Dossenheim (DE).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)*

— *of inventorship (Rule 4.17(iv)) for US only*

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NOVEL MAMMALIAN NUCLEAR RECEPTOR L66 AND METHODS OF USE

(57) **Abstract:** The present invention relates to a novel nuclear receptor called "L66" or also FXR-β a homologue of the FXR-α, a prototypical type 2 nuclear receptor. The invention also relates to the isolated nucleic acid sequence of L66 and the isolated protein thereof. The invention further relates to processes for isolating and/or producing the nucleic acid or the protein as well as methods of use of the receptor L66.

WO 02/22817 A2

Title:

NOVEL MAMMALIAN NUCLEAR RECEPTOR L66 AND METHODS OF USE

BACKGROUND OF THE INVENTION

Multicellular organisms are dependent on advanced mechanisms of information transfer between cells and body compartments. The information that is transmitted can be highly complex and can result in the alteration of genetic programs involved in cellular differentiation, proliferation, or reproduction. The signals, or hormones, are often simple molecules, such as peptides, fatty acid, or cholesterol derivatives.

Many of these signals produce their effects by ultimately changing the transcription of specific genes. One well-studied group of proteins that mediate a cell's response to a variety of signals is the family of transcription factors known as nuclear receptors, hereinafter referred to often as "NR". Members of this group include receptors for steroid hormones, vitamin D, ecdysone, cis and trans retinoic acid, thyroid hormone, bile acids, cholesterol-derivatives, fatty acids (and other peroxisomal proliferators), as

well as so-called orphan receptors, proteins that are structurally similar to other members of this group, but for which no ligands are known (Escriva, H. et al., Ligand binding was acquired during evolution of nuclear receptors, PNAS, 94, 6803 – 6808, 1997). Orphan receptors may be indicative of unknown signaling pathways in the cell or may be nuclear receptors that function without ligand activation. The activation of transcription by some of these orphan receptors may occur in the absence of an exogenous ligand and/or through signal transduction pathways originating from the cell surface (Mangelsdorf, D. J. et al., The nuclear receptor superfamily: the second decade, Cell 83, 835-839, 1995).

In general, three functional domains have been defined in NRs. An amino terminal domain is believed to have some regulatory function. A DNA-binding domain hereinafter referred to as "DBD" usually comprises two zinc finger elements and recognizes a specific Hormone Responsive Element hereinafter referred to as "HRE" within the promoters of responsive genes. Specific amino acid residues in the "DBD" have been shown to confer DNA sequence binding specificity (Schena, M. & Yamamoto, K.R., Mammalian Glucocorticoid Receptor Derivatives Enhance Transcription in Yeast, Science, 241:965-967, 1988). A Ligand-binding-domain hereinafter referred to as "LBD" is at the carboxy-terminal region of known NRs. In the absence of hormone, the LBD appears to interfere with the interaction of the DBD with its HRE. Hormone binding seems to result in a conformational change in the NR and thus opens this interference (Brzozowski et al., Molecular basis of agonism and antagonism in the oestogen receptor, Nature, 389, 753 – 758, 1997; Wagner et al., A structural role for hormone in the thyroid hormone receptor, Nature, 378, 690 – 697. 1995). A NR without the HBD constitutively activates transcription but at a low level.

Both the amino-terminal domain and the LBD appear to have transcription activation functions hereinafter referred to as "TAF". Acidic residues in the amino-terminal domains of some nuclear receptors may be important for these transcription factors to interact with RNA polymerase. TAF activity may be dependent on interactions with other protein factors or nuclear components (Diamond et al., Transcription Factor Interactions: Selectors of Positive or Negative Regulation from a Single DNA Element, Science, 249:1266-1272 , 1990). Certain oncoproteins (e.g., c-Jun and c-Fos) can show synergistic or antagonistic activity with glucocorticoid receptors (GR)

in transfected cells. Furthermore, the receptors for estrogen and vitamins A and D, and fatty acids have been shown to interact, either physically or functionally, with the Jun and Fos components of AP-1 in the transactivation of steroid- or AP-1 regulated genes.

Coactivators or transcriptional activators are proposed to bridge between sequence specific transcription factors, the basal transcription machinery and in addition to influence the chromatin structure of a target cell. Several proteins like SRC-1, ACTR, and Grip1 interact with NRs in a ligand enhanced manner (Heery et al., A signature motif in transcriptional coactivators mediates binding to nuclear receptors, Nature, 387, 733 – 736; Heinzel et al., A complex containing N-CoR, mSin3 and histone deacetylase mediates transcriptional repression, Nature 387, 43 – 47, 1997). Furthermore, the physical interaction with negative receptor-interacting proteins or corepressors has been demonstrated (Xu et al., Coactivator and Corepressor complexes in nuclear receptor function, Curr Opin Genet Dev, 9 (2), 140 – 147, 1999).

Nuclear receptor modulators like steroid hormones affect the growth and function of specific cells by binding to intracellular receptors and forming nuclear receptor-ligand complexes. Nuclear receptor-hormone complexes then interact with a hormone response element (HRE) in the control region of specific genes and alter specific gene expression.

Over the past decade, new members of the nuclear hormone gene family have been identified that lack known ligands. These orphan receptors can be used to uncover signaling molecules that regulate yet unidentified physiological networks. Some of these orphan receptors are constitutively active in transactivate target genes without the need to interact with a ligand (Mangelsdorfet al., 1995).

Farnesoid X Receptor alpha (hereinafter FXR-α) is a prototypical type 2 nuclear receptor (US Pat. 6,005,086) which activates genes upon binding to promoter region of target genes in a heterodimeric fashion with Retinoid X Receptor (hereinafter RXR, Forman et al., Cell, 81, 687-93, 1995).. The relevant physiological ligands of FXR-α seem to be bile acids. The most potent is chenodeoxycholic acid, which regulates the expression of several genes that participate in bile acid homeostisis. Farnesoid, originally

described to activate the rat ortholog at high concentration does not activate the human or mouse receptor. It is highly expressed in the liver, intestine and kidney. Like LXR-α FXR-α is involved in intracrine signaling.

The relevant physiological ligands of NR1H4 (as FXR-α is also called) seem to be bile acids (Makishima et al., Science, 284, 1362-65, 1999; Parks et al., Science, 284, 1365-68, 1999). The most potent is chenodeoxycholic acid, which regulates the expression of several genes that participate in bile acid homeostasis.

Consequently, FXR-α is proposed to be a nuclear bile acid sensor. As a result, it modulates both, the synthetic output of bile acids in the liver and their recycling in the intestine (by regulating bile acid binding proteins). Its is also activated by retoinic acid and TTNPB at supraphisiological concentration. Furthermore, it regulates the conversion of dietry cholesterol into bile acids by regulation the metabolizing genes like CYP7-α. This is a feed back regulation since the receptor is activated by bile acids.

Through its regulatory function in cholesterol and bile acid metabolism an FXR-α homologue could serve as a target for cholesterol lowering drugs and exert beneficial effects in diseases like artheriosclerosis and other metabolic disorders.

It was thus an object of the present invention to provide for a novel nuclear receptor. In a preferred embodiment of the invention it was an object to provide for a homologue of FXR-α. It was an object of the present invention to provide for means of producing this receptor as well as means of screening for agonists and antagonists to the receptor. Further objects of the invention are outlined below.

SUMMARY OF THE INVENTION

The present invention provides, *inter alia*, a novel nuclear receptor protein. In a preferred embodiment of the invention a novel FXR-α homologue is provided for. Also provided is the nucleic acid sequence encoding this novel nuclear receptor protein, as well as compounds and methods for using this protein and its nucleic acid sequence.

The present invention provides a novel proteins, nucleic acids, and methods useful for developing and identifying compounds for the treatment of such diseases and disorders as metabolic disorders, immunological indications, hormonal dysfunctions, neurosystemic diseases and in preferred embodiments, high cholesterol and artheriosclerosis as well as other metabolic disorders.

Identified and disclosed herein is the protein sequence for a novel nuclear receptor and the nucleic acid sequence encoding this nuclear receptor L66, which we call the L66 nuclear receptor (or simply "L66") receptor, or also FXR-β.

The importance of this discovery is manifested in the effects of FXR-β to modulate genes involved in cellular functions like regulation of metabolism and cell homeostasis, cell proliferation and differentiation, pathological cellular aberrations, or cellular defense mechanisms including tumor development, *i.e.* cancer.

Thus, this L66 protein is useful for screening for L66 agonists and antagonist activity for controlling these conditions.

In one aspect of the present invention, we provide isolated nucleic acid sequences for a novel *receptor*, the L66 receptor. In particular, we provide the cDNA sequences, protein sequences as well as the genomic sequences encoding the human L66 receptor, as well as the cDNA sequence, protein sequence and genomic sequence of the Mus musculus (mouse) receptor.

These nucleic acid sequences have a variety of uses. For example, they are useful for making vectors and for transforming cells, both of which are ultimately useful for production of the L66 protein.

They are also useful as scientific research tools for developing nucleic acid probes for determining L66 expression levels, e.g., to identify diseased or otherwise abnormal states. They are useful for developing analytical tools such as anti sense oligonucleotides for selectively inhibiting expression of the L66 gene to det rmine physiological responses.

In another aspect of the present invention, we provide a homogenous composition comprising the L66 protein. The protein is useful for screening drugs for agonist and antagonist activity, and, therefore, for screening for drugs useful in regulating physiological responses associated with L66. Specifically, antagonists to the L66 receptor could be used to treat diseases and disorders as metabolic disorders, immunological indications, hormonal dysfunctions, neurosystemic diseases and in preferred embodiments, high cholesterol and artheriosclerosis as well as other metabolic disorders, whereas agonists could be used for the treatment of these conditions. The proteins are also useful for developing antibodies for detection of the protein.

Flowing from the foregoing are a number of other aspects of the invention, including (a) vectors, such as plasmids, comprising the L66 nuclear receptor nucleic acid sequence that may further comprise additional regulatory elements, *e.g.*, promoters, (b) transformed cells that express the L66, (c) nucleic acid probes, (d) antisense oligonucleotides, (e) agonists, (f) antagonists, and (g) transgenic mammals. Further aspects of the invention comprise methods for making and using the foregoing compounds and compositions.

The foregoing merely summarizes certain aspects of the present invention and is not intended, nor should it be construed, to limit the invention in any manner. All patents and other publications recited herein are hereby incorporated by reference in their entirety.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

THE L66 PROTEIN AND NUCLEIC ACID:

The present invention comprises a novel member of the nuclear receptor superfamily which the inventors herein refer to as L66. Particularly preferred embodiments of the L66 receptor are those having an amino acid sequence substantially the same as SEQ ID NO. 3, 24 and/or 19. Examination of the amino acid sequence confirms that the present protein is indeed a member of the nuclear receptor family (see also US 6,005,086) which is closely related to FXR (see also figure 3 for the domain

composition). The carboxy-terminal ligand binding domain "LBD" of L66 is a complex region encoding subdomains for ligand binding, often dimerization and transcriptional activation.

The nucleic acids claimed above may be present in various forms, i.e. as an RNA, DNA, cDNA or as genomic DNA.

As used herein, if reference to L66, the L66 receptor, the nuclear receptor L66 or the L66 nuclear receptor is made it is meant as a reference to any protein having an amino acid sequence substantially the same as SEQ ID NO.: 3, 24 and/or 19. The present invention also comprises nucleic acid sequences encoding the L66 receptor, which nucleic acid sequences are substantially the same as SEQ ID NO. 1and SEQ ID NO. 17 (cDNAs) or splice variants thereof 4, 6, 8, 10, 12, 22, their sequence complements SEQ ID NO. 2 and 18 or complements of said splice variants 5, 7, 9, 11, 13, 23. SEQ ID NO 4encodes the human cDNA L66 receptor and is a preferred embodiment. SEQ OD NO. 17 (cDNA) encodes the mouse receptor (Mus musculus). SEQ ID NO. 20 represents the genomic sequence of L66 locus from the mouse (Mus musculus) (see also Fig. 15)in one embodiment, the nucleic acid sequence according to the invention comprises the sequence according to SEQ ID NO. 1, 4, 6, 8, 10, 12, 17, 20(4 to 12 representing various splice variants of the human L66) and/or 22 or portions thereof, in a preferred embodiment nucleic acid sequence according to the invention consists of the sequence according to SEQ ID NO. 1, 4, 6, 8, 10 (4 to 10 representing various splice variants of the human L66) and/or 17 or portions thereof.

Herein the "complement" refers to the complementary strand of the nucleic acid according to the invention, thus the strand that would hybridize to the nucleic acid according to the invention. In accordance with standard biological terminology all DNA sequences herein are however written in 5'-3' orientation, thus the if a complement is mentioned (see also figures) it is actually a "reverse" complement (as also stated in the figures). For simplification purposes they may however sometimes be referred to simply as "complements".

As used herein, a protein "having an amino acid sequence substantially the same or similar as SEQ ID NO x" (where "x" is the number of one of the protein sequences recited in the Sequence Listing) means a protein whose amino acid sequence is the same as SEQ ID NO x or differs only in a way such that at least 50% of the residues compared in a sequence alignment with SEQ ID NO. x are identical, preferably 75% of the residues are identical, even more preferably 95% of the residues are identical and most preferably at least 98% of the residues are identical.

Those skilled in the art will appreciate that conservative substitutions of amino acids can be made without significantly diminishing the protein's affinity for interacting proteins, DNA binding sites, L66 receptor modulators, *e.g.* small molecular hydrophobic compounds, or RNA.

Other substitutions may be made that increase the protein's affinity for these compounds. Making and identifying such proteins is a routine matter given the teachings herein, and can be accomplished, for example, by altering the nucleic acid sequence encoding the protein (as disclosed herein), inserting it into a vector, transforming a cell, expressing the nucleic acid sequence, and measuring the binding affinity of the resulting protein, all as taught herein.

As used herein the term "a molecule having a nucleotide sequence substantially the same as SEQ ID NO y" (wherein "y" is the number of one of the protein-encoding nucleotide sequences listed in the Sequence Listing) means a nucleic acid encoding a protein "having an amino acid sequence substantially the same as SEQ ID NO y+l" (wherein "y+l" is the number of the amino acid sequence for which nucleotide sequence "y" codes) as defined above. This definition is intended to encompass natural allelic variations in the L66 sequence. Cloned nucleic acid provided by the present invention may encode L66 protein of any species of origin, including (but not limited to), for example, mouse, rat, rabbit, cat, dog, primate, and human.

Preferably, the nucleic acid provided by the invention encodes L66 receptors of mammalian, preferably mouse and most preferably human origin.

Preferably, the L66 receptors proteins provided by the invention are of mammalian, more preferably mouse and most preferably human origin.

The inventors have found (see figures and examples) that the L66 gene in humans is expressed primarily in testis. In mice however expression may be observed also in other tissues.

IDENTIFICATION OF VARIANTS AND HOMOLOGUES AS WELL AS USE OF PROBES:

Nucleic acid hybridization probes provided by the invention are nucleic acids consisting essentially of the nucleotide sequences complementary to any sequence depicted in SEQ ID NO. 1, 4, 6, 8, 10, 12, 17, 20, 22, 2, 5, 7, 9, 11, 13, 18, 21 and/or 23 or a part thereof and that are effective in nucleic acid hybridization

Nucleic acid hybridization probes provided by the invention are nucleic acids capable of detecting i.e. hybridizing to the gene encoding the polypeptides according to SEQ ID NO. 3, 24 and/or 19.

Nucleic acid probes are useful for detecting L66 gene expression in cells and tissues using techniques well-known in the art, including, but not limited to, Northern blot hybridization, in situ hybridization, and Southern hybridization to reverse transcriptase - polymerase chain reaction product DNAs. The probes provided by the present invention, including oligonucleotide probes derived therefrom, are also useful for Southern hybridization of mammalian, preferably human, genomic DNA for screening for restriction fragment length polymorphism (RFLP) associated with certain genetic disorders. As used herein, the term complementary means a nucleic acid having a sequence that is sufficiently complementary in the Watson-Crick sense to a target nucleic acid to bind to the target under physiological conditions or experimental conditions those skilled in the art routinely use when employing probes.

It is understood in the art that a nucleic acid sequence will hybridize with a complementary nucleic acid sequence under high stringent conditions as defined

herein, even though some mismatches may be present. Such closely matched, but not perfectly complementary sequences are also encompassed by the present invention. For example, differences may occur through genetic code degeneracy, or by naturally occurring or man made mutations and such mismatched sequences would still be encompassed by the present claimed invention.

Preferably, the nucleotide sequence of the nuclear receptor L66 (SEQ ID NO. 1, 4, 17, and/or 22 or splice variants thereof) and/or their complements can be used to derive oligonucleotide fragments (probes) of various length. If the probe is used to detect a human L66 sequence most preferably a complement of SEQ ID NO. 1 and/or 4 or its complement is used. If the probe is supposed to detect a mouse or a rodent sequence probes complementary to SEQ ID NO. 17 and/or 22, or their respective complements are preferred. Stretches of 17 to 30 nucleotides are used frequently but depending on the screening parameters longer sequences as 40, 50, 100, 150 up to the full length of the sequence may be used. Those probes can be synthesized chemically and are obtained readily from commercial oligonucleotide providers. Chemical synthesis has improved over the years and chemical synthesis of oligonucleotides as long as 100-200 bases is possible. The field might advance further to allow chemical synthesis of even longer fragments. Alternatively, probes can also be obtained by biochemical *de novo* synthesis of single stranded DNA. In this case the nucleotide sequence of the nuclear receptor L66 or its complement (see figures) serve as a template and the corresponding complementary strand is synthesized. A variety of standard techniques such as nick translation or primer extension from specific primers or short random oligonucleotides can be used to synthesize the probe (Molecular Cloning: A Laboratory Manual (3 Volume Set) by Joseph Sambrook, David W. Russell, Joe Sambrook, 2100 pages 3rd edition (January 15, 2001; . Molecular cloning: a laboratory manual. Cold Spring Harbor Press, Cold Spring Harbor, 1989)). Nucleic acid reproduction technologies exemplified by the polymerase chain reaction (Saiki, R.K. et *al.* Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239, 487-491 (1988)) are commonly applied to synthesize probes. In the case of techniques using specific primers the nucleic acid sequence of the nuclear receptor L66 or its complement is not only used as a template in the biochemical reaction but also to derive the specific primers which are needed to prime the reaction.

In some cases one might also consider to use the nucleic acid sequence of the cofactor or its complement as a template to synthesize an RNA probe. A promoter sequence for a DNA-dependent RNA polymerase has to be introduced at the 5'-end of sequence. As an example this can be done by cloning the sequence into a vector which carries the respective promoter sequence. It is also possible to introduce the needed sequence by synthesizing a primer with the needed promoter in the form of a 5' "tail". The chemical synthesis of a RNA probe is another option.

Appropriate means are available to detect the event of a hybridization. There is a wide variety of labels and detection systems, e.g. radioactive isotopes, fluorescent, or chemiluminescent molecules which can be linked to the probe. Furthermore, there are methods of introducing haptens which can be detected by antibodies or other ligands such as the avidin/biotin high affinity binding system.

Hybridization can take place in solution or on solid phase or in combinations of the two, e.g. hybridization in solution and subsequent capture of the hybridization product onto a solid phase by immobilized antibodies or by ligand coated magnetic beads.

Hybridization probes act by forming selectively duplex molecules with complementary stretches of a sequence of a gene or a cDNA. The selectivity of the process can be controlled by varying the conditions of hybridization. To select sequences which are identical highly homologous to the sequence of interest stringent conditions for the hybridization will be used, e.g. low salt in the range of 0.02 M to 0.15 M salt and/or high temperatures in the range from 50°C degrees centigrade to 70°C degrees centigrade. Stringency can be further improved by the addition of formamide to the hybridization solution. The use of stringent conditions which means that only little mismatch or a complete match will lead to a hybridization product would be used to isolate closely related members of the same gene family. Thus, as used herein stringent hybridization conditions are those where between 0.02 M to 0.15 M salt and/or high temperatures in the range from 50°C degrees centigrade to 70°C degrees centigrade are applied.

The use of highly stringent conditions or conditions of "high stringency" means that only very little mismatch or a complete match which lead to a hybridization product would be used to isolate very closely related members of the same gene family. Thus, as used herein highly stringent hybridization conditions are those where between 0.02 – 0.3 M salt and 65°C degrees centigrade are applied for about 5 to 18 hours of hybridization time and additionally, the sample filters are washed twice for about 15 minutes each at between 60°C – 65°C degrees centigrade, wherein the first washing fluid contains about 0.1 M salt (NaCl and/or Sodium Citrate) and the second contains only about 0.02 M salt (NaCl and/or Sodium Citrate). In a preferred embodiment the following conditions are considered to be highly stringent:

Hybridisation in a buffer containing 2 x SSC (0.03 M Sodium Citrate, 0.3 M NaCl) at 65°C – 68°C degrees centigrade for 12 hours, followed by a washing step for 15 minutes in 0.5 x SSC, 0.1% SDS, and a washing step for 15 minutes at 65°C degrees centigrade in 0.1 x SSC, 0.1% SDS.

Less stringent hybridization conditions, e.g. 0.15 M salt - 1 M salt and/or temperatures from 22°C degrees centigrade to 56°C degrees centigrade are applied in order to detect functionally equivalent genes in the same species or for orthologous sequences from other species.

Unspecific hybridization products are removed by washing the reaction products repeatedly in 2 x SSC solution and increasing the temperature.

DEGENERATE PCR AND CLONING OF HOMOLOGUES

The nucleotide sequence of the nuclear receptor L66 or its complement can be used to design primers for a polymerase chain reaction. Due to the degeneracy of the genetic code the respective amino acid sequence is used to design oligonucleotides in which varying bases coding for the same amino acid are included. Numerous design rules for degenerate primers have been published (Compton et al, 1990). As in hybridization there are a number of factors known to vary the stringency of the PCR. The most important parameter is the annealing temperature. To allow annealing of primers with imperfect matches annealing temperatures are often much

lower than the standard annealing temperature of 55°C, e.g. 35°C to 52°C degrees can be chosen. PCR reaction products can be cloned. Either th PCR product is cloned directly, with reagents and protocols from commercial manufacturers (*e.g.* from Invitrogen, San Diego, USA). Alternatively, restriction sites can be introduced intro the PCR product via a 5'-tail of the PCR primers and used for cloning. Primers for the amplification of the entire or partial pieces of the L66 gene or mRNA, or for reverse transcription may be designed making use of the sequences according to the invention, i.e. those depicted in the figures below.

GENETIC VARIANTS

Fragments from the nucleotide sequence of the nuclear receptor L66 (SEQ ID NO. 1, 4 or the mouse, ,i.e. 17 or 22) or their complements can be used to cover the whole sequence with overlapping sets of PCR primers. Also the genomic sequences may be used (see figures for sequences). These primers are used to produce PCR products using genomic DNA from a human diversity panel of healthy individuals or genomic DNA from individuals which are phenotypically conspicuous. Also the genomic sequences may be used, *i.e.* that of the human clone as deposited by the applicant (deposit number DSM 14483) or that of the mouse according to SEQ ID NO. 20 (or the complement thereof). The PCR products can be screened for polymorphisms, for example by denaturing gradient gel electrophoresis, binding to proteins detecting mismatches or cleaving heteroduplices or by denaturing high-performance liquid chromatography. Products which display mutations need to be sequenced to identify the nature of the mutation. Alternatively, PCR products can be sequenced directly omitting the mutation screening step to identify genetic polymorphisms. If genetic variants are identified and are associated with a discrete phenotype, these genetic variations can be included in diagnostic assays. The normal variation of the human population is of interest in designing screening assays as some variants might interact better or worse with a respective lead substance (a pharmacodynamic application). Polymorphisms or mutations which can be correlated to phenotypic outcome are a tool to extend the knowledge and the commercial applicability of the nucleotide sequence of the nuclear receptor L66 or its complement or their gene product, as variants might have a slightly different molecular behavior or desired properties. Disease-causing mutations or

polymorphisms allow the replacement of this disease inducing gene copy with a wild-type copy by means of gene therapy approaches and/or the modulation of th activity of the gene product by drugs.

PREPARATION OF POLYNUCLEOTIDES:

DNA which encodes receptor L66 may be obtained, in view of the instant disclosure, by chemical synthesis, by screening reverse transcripts of mRNA from appropriate cells or cell line cultures, by screening genomic libraries from appropriate cells, or by combinations of these procedures, as illustrated below.

Screening of mRNA or genomic DNA may be carried out with oligonucleotide probes generated from the L66 nucleotide sequences information provided herein.

Probes may be labeled with a detectable group such as a fluorescent group, a radioactive atom or a chemiluminescent group in accordance with known procedures and used in conventional hybridization assays, as described in greater detail in the Examples below. Alternatively, the L66 nucleotide sequence may be obtained by use of the polymerase chain reaction (PCR) procedure, with the PCR oligonucleotide primers being produced from the L66 nucleotide sequences provided herein.

Upon purification or synthesis, the nucleic acid according to the invention may be labeled, e.g. for use as a probe.

As single and differential labeling agents and methods, any agents and methods which are known in the art can be used. For example, single and differential labels may consist of the group comprising enzymes such as β-galactosidase, alkaline phosphatase and peroxidase, enzyme substrates, coenzymes, dyes, chromophores, fluorescent, chemiluminescent and bioluminescent labels such as FITC, Cy5, Cy5.5, Cy7, Texas-Red and IRD40(Chen et al. (1993), J. Chromatog. A 652: 355-360 and Kambara et al. (1992), Electrophoresis 13: 542-546), ligands or haptens such as biotin, and radioactive isotopes such as ^{3}H, ^{35}S, ^{32}P ^{125}I and ^{14}C.

EXPRESSION OF THE L66 PROTEIN/POLYPETIDE:

The nuclear receptor L66 nucleic acid or polypeptide may be synthesized in host cells transformed with a recombinant expression construct comprising a nucleic acid encoding the nuclear receptor L66.

Such a recombinant expression construct can also be comprised of a vector that is a replicable DNA construct.

Amplification vectors do not require expression control domains. All that is needed is the ability to replicate in a host, usually conferred by an origin of replication, and a selection gene to facilitate recognition of transformants. See, Sambrook et al., Molecular Cloning: A Laboratory Manual (2nd Edition, Cold Spring Harbor Press, New York, 1989).

An expression vector comprises a polynucleotide operatively linked to a prokaryotic promoter. Alternatively, an expression vector is a polynucleotide operatively linked to an enhancer-promoter that is a eukaryotic promoter, and the expression vector further has a polyadenylation signal that is positioned 3' of the carboxy-terminal amino acid and within a transcriptional unit of the encoded polypeptide. A promoter is a region of a DNA molecule typically within about 500 nucleotide pairs in front of (upstream of) the point at which transcription begins (*i.e.*, a transcription start site). In general, a vector contains a replicon and control sequences which are derived from species compatible with the host cell. The vector ordinarily carries a replication site, as well as marking sequences which are capable of providing phenotypic selection in transformed cells.

Another type of discrete transcription regulatory sequence element is an enhancer. An enhancer provides specificity of time, location and expression level for a particular encoding region (e.g., gene). A major function of an enhancer is to increase the level of transcription of a coding sequence in a cell.

As used herein, the phrase "enhancer-promoter" means a composite unit that contains both enhancer and promoter elements. An enhancer-promoter is operatively linked to a coding sequence that encodes at least one gene product.

An enhancer-promoter used in a vector construct of the present invention may be any enhancer-promoter that drives expression in a prokaryotic or eukaryotic cell to be transformed/transfected.

A coding sequence of an expression vector is operatively linked to a transcription terminating region. RNA polymerase transcribes an encoding DNA sequence through a site where polyadenylation occurs.

An expression vector comprises a polynucleotide that encodes a nuclear receptor L66 polypeptide. Such a polynucleotide is meant to include a sequence of nucleotide bases encoding a nuclear receptor L66 polypeptide sufficient in length to distinguish said segment from a polynucleotide segment encoding a non- nuclear receptor L66 polypeptide.

A polypeptide of the invention may also encode biologically functional polypeptides or peptides which have variant amino acid sequences, such as with changes selected based on considerations such as the relative hydropathic score of the amino acids being exchanged.

These variant sequences are those isolated from natural sources or induced in the sequences disclosed herein using a mutagenic procedure such as site-directed mutagenesis.

Furthermore, an expression vector of the present invention may contain regulatory elements for optimized translation of the polypeptide in prokaryotic or eukaryotic systems. This sequences are operatively located around the transcription start site and are most likely similar to ribosome recognition sites like prokaryotic ribosome binding sites (RBS) or eukaryotic Kozak sequences as known in the art (Kozak M., Initiation of translation in prokaryotes and eukaryotes. *Gene* 234, 187-208 (1999).

An expression vector of the present invention is useful both as a means for preparing quantities of the nuclear receptor L66 polypeptide-encoding DNA itself, and as a means for preparing the encoded nuclear receptor L66 polypeptide and peptides. It is contemplated that where nuclear receptor L66 polypeptides of the invention are made by recombinant means, one may employ either prokaryotic or eukaryotic expression vectors as shuttle systems.

Where expression of recombinant nuclear receptor L66 polypeptides is desired and a eukaryotic host is contemplated, it is most desirable to employ a vector such as a plasmid, that incorporates a eukaryotic origin of replication. Additionally, for the purposes of expression in eukaryotic systems, one desires to position the nuclear receptor L66 encoding sequence or if desired parts thereof (SEQ ID NO. 1, 4, 6, 8, 10, 12, 17, and/or 22) adjacent to and under the control of an effective eukaryotic promoter. To bring a coding sequence under control of a promoter, whether it is eukaryotic or prokaryotic, what is generally needed is to position the 5' end of the translation initiation site of the proper translational reading frame of the polypeptide between about 1 and about 2000 nucleotides 3' of or downstream with respect to the promoter chosen.

Furthermore, where eukaryotic expression is anticipated, one would typically desire to incorporate into the transcriptional unit which includes the nuclear receptor L66 polypeptide, an appropriate polyadenylation side.

The invention provides homogeneous compositions of mammalian nuclear receptor L66 polypeptide produced by transformed prokaryotic or eukaryotic cells as provided herein. Such homogeneous compositions are intended to be comprised of mammalian nuclear receptor L66 protein that comprises at least 90% of the protein in such homogenous composition. The invention also provides membrane preparation from cells expressing mammalian nuclear receptor L66 polypeptide as the result of transformation with a recombinant expression construct, as described here.

Within the scope of the present invention the terms recombinant protein or coding sequence both also include tagged versions of the protein depicted in SEQ ID NO. 3, 19 and/or 22, and/or encoded by the nucleic acids according to the invention and

fusion proteins of said proteins or parts thereof such as splice variants with any other recombinant protein. Tagged versions here means that small epitopes of 3-20 amino acids are added to the original protein by extending the coding sequence either at the 5'or the 3'terminus leading to N-terminal or C-terminal extended proteins respectively, or that such small epitopes are included elsewhere in the protein. The same applies for fusion proteins where the added sequences are coding for longer proteins, varying between 2 and 100 kDa. Tags and fusion proteins are usually used to facilitate purification of recombinant proteins by specific antibodies or affinity matrices or to increase solubility of recombinant proteins within the expression host. Fusion proteins are also of major use as essential parts of yeast two hybrid screens for interaction partners of recombinant proteins.

Tags used in the scope of the present invention may include but are not limited to the following: EEF (alpha Tubulin), B-tag (QYPALT), E tag (GAPVPYPDPLEPR) c-myc Tag (EQKLISEEDL), Flag epitope (DYKDDDDK, HA tag (YPYDVPDYA), 6 or 10 x His Tag, HSV (QPELAPEDPED), Pk-Tag (GKPIPNPLLGLDST), protein C (EDQVDPRLIDGK), T7 (MASMTGGQQMG), VSV-G (YTDIEMNRLGK), Fusion proteines may include Thioredoxin, Glutathiontransferase (GST), Maltose binding Protein (MBP), Cellulose Binding protein (CBD), chitin binding protein, ubiquitin, the Fc part of Immunoglobulins, and the IgG binding domain of *Staphylococcus aureus* protein A. These examples of course are illustrative and not limiting and the standard amino acid one letter code was used above.

For expression of recombinant proteins in living cells or organisms, vector constructs harboring recombinant L66 nuclear receptor as set forth in SEQ ID NO. 1 or 17 are transformed or transfected into appropriate host cells. Preferably, a recombinant host cell of the present invention is transfected with a polynucleotide of SEQ ID NO. 1, 4, 22 or 17.

Means of transforming or transfecting cells with exogenous polynucleotide such as DNA molecules are well known in the art and include techniques such as calcium-phosphate- or DEAE-dextran-mediated transfection, protoplast fusion, electroporation, liposome mediated transfection, direct microinjection and virus infection (Sambrook et al., 1989).

The most frequently applied technique for transformation of prokaryotic cells is transformation of bacterial cells after treatment with calciumchloride to increase permeability (Dagert & Ehrlich, 1979), but a variety of other methods is also available for one skilled in the art.

The most widely used method for transfection of eukaryotic cells is transfection mediated by either calcium phosphate or DEAE-dextran. Although the mechanism remains obscure, it is believed that the transfected DNA enters the cytoplasm of the cell by endocytosis and is transported to the nucleus. Depending on the cell type, up to 90% of a population of cultured cells may be transfected at any one time. Because of its high efficiency, transfection mediated by calcium phosphate or DEAE-dextran is the method of choice for studies requiring transient expression of the foreign nucleic acid in large numbers of cells. Calcium phosphate-mediated transfection is also used to establish cell lines that integrate copies of the foreign DNA, which are usually arranged in head-to-tail tandem arrays into the host cell genome.

In the protoplast fusion method, protoplasts derived from bacteria carrying high numbers of copies of a plasmid of interest are mixed directly with cultured mammalian cells. After fusion of the cell membranes (usually with polyethylene glycol), the contents of the bacterium are delivered into the cytoplasm of the mammalian cells and the plasmid DNA is transported to the nucleus. Protoplast fusion is not as efficient as transfection for many of the cell lines that are commonly used for transient expression assays, but it is useful for cell lines in which endocytosis of DNA occurs inefficiently. Protoplast fusion frequently yields multiple copies of the plasmid DNA tandemly integrated into the host chromosome.

The application of brief, high-voltage electric pulses to a variety of mammalian and plant cells leads to the formation of nanometer sized pores in the plasma membrane. DNA is taken directly into the cell cytoplasm either through these pores or as a consequence of the redistribution of membrane components that accompanies closure of the pores. Electroporation may be extremely efficient and may be used both for transient expression of cloned genes and for establishment of cell lines that carry integrated copies of the gene of interest. Electroporation, in contrast to calcium

phosphate-mediated transfection and protoplast fusion, frequently gives rise to cell lines that carry one, or at most a few, integrated copies of the foreign DNA.

Liposome transfection involves encapsulation of DNA and RNA within liposomes, followed by fusion of the liposomes with the cell membrane. The mechanism of how DNA is delivered into the cell is unclear but transfection efficiencies may be as high as 90%.

Direct microinjection of a DNA molecule into nuclei has the advantage of not exposing DNA to cellular compartments such as low-pH endosomes. Microinjection is therefore used primarily as a method to establish lines of cells that carry integrated copies of the DNA of interest.

The use of adenovirus as a vector for cell transfection is well known in the art. Adenovirus vector-mediated cell transfection has been reported for various cells (Stratford-Perricaudet et al., 1992).

A transfected cell may be prokaryotic or eukaryotic, transfection may be transient or stable. Where it is of interest to produce a full length human or mouse L66 protein, cultured mammalian mouse, or human cells are of particular interest.

In another aspect, the recombinant host cells of the present invention are prokaryotic host cells. In addition to prokaryotes, eukaryotic microbes, such as yeast may also be used illustrative examples for suitable cells and organisms for expression of recombinant proteins are belonging to but not limited to the following examples:

- Insect cells, such as Drosophila Sf21, SF9 cells or others, Expression strains of Escherichia coli, such as XL1 blue, BRL21, M15, Saccharomyces cerevisiae, Schizosaccharomyces pombe, *Hansenlua polymorpha and Pichia pastoris* strains, immortalized mammalian cell lines such as AtT-20, VERO and HeLa cells, Chinese hamster ovary (CHO) cell lines, and W138, BHK, COSM6, COS-7, 293 and MDCK cells, BHK-21 cells, Att 20HeLa cells, HeK 294, T47 D cells and others.

Expression of recombinant proteins within the scope of this invention can also be performed *in vitro*. This may occur by a two step procedure, thereby producing first

mRNA by in vitro transcription of an apt polynucleotide construct followed by in vitro translation with convenient cellular extracts. These cellular extracts may be reticulocyte lysates but are not limited to this type. In vitro transcription may be performed by T7 or SP6 DNA polymerase or any other RNA polymerase which can recognize *per se* or with the help of accessory factors the promoter sequence contained in the recombinant DNA construct of choice. Alternatively one of the recently made available one step coupled transcription/translation systems may be used for in vitro translation of DNA coding for the proteins of this invention, e.g. from Roche Molecular Biochemicals. One illustrative but not limiting example for such a system is the TNT® T7 Quick System by Promega.

Expression of recombinant proteins in transfected cell may occur constitutively or upon induction. Procedures depend on the Cell/vector combination used and are well known in the art. In all cases, transfected cells are maintained for a period of time sufficient for expression of the recombinant L66 nuclear receptor protein. A suitable maintenance time depends strongly on the cell type and organism used and is easily ascertainable by one skilled in the art. Typically, maintenance time is from about 2 hours to about 14 days. For the same reasons and for sake of protein stability and solubility incubation temperatures during maintenance time may vary from 20°C to 42 °C.

Recombinant proteins are recovered or collected either from the transfected cells or the medium in which those cells are cultured. Recovery comprises cell disruption, isolation and purification of the recombinant protein. Isolation and purification techniques for polypeptides are well-known in the art and include such procedures as precipitation, filtration, chromatography, electrophoresis and the like.

In a preferred embodiment, purification includes but is not limited to affinity purification of tagged or non-tagged recombinant proteins. This is a well established robust technique easily adapted to any tagged protein by one skilled in the art. For affinity purification of tagged proteins, small molecules such as gluthathione, maltose or chitin, specific proteins such as the IgG binding domain of Staphylococcus aureus protein A, antibodies or specific chelates which bind with high affinity to the tag of the recombinant protein are employed. For affinity purification of non-tagged proteins

specific monoclonal or polyclonal antibodies, which were raised against said protein, can be used. Alternatively immobilized specific interactors of said protein may be employed for affinity purification. Interactors include native or recombinant proteins as well as native or artificial specific low molecular weight ligands.

CHEMICAL SYNTHESIS OF THE POLYPEPTIDE ACCORDING TO THE INVENTION:

Alternatively, the protein itself may be produced using chemical methods to synthesize any of the amino acid sequences according to the invention or that is encoded by the nucleotide sequences according to the invention (SEQ ID NO. 1, 4, 17 or 22) and/or a portion thereof and/or splice variants thereof. For example, peptide synthesis can be performed using conventional Merrifield solid phase f-Moc or t-Boc chemistry or various solid-phase techniques (Roberge, J. Y. et al. (1995) Science 269: 202-204) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer). The newly synthesized peptide(s) may be substantially purified by preparative high performance liquid chromatography (e.g., Creighton, T. (1983) Proteins, Structures and Molecular Principles, WH Freeman and Co., New York, N.Y.). The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (*e.g.,* the Edman degradation procedure; Creighton, supra). Additionally, the amino acid sequences according to the invention, *i.e.* SEQ ID NO. 3, 24 or 19 or the sequence that is encoded by SEQ ID NO. 1, 4, 17 or 22 or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

SCREENING ASSAYS

The invention also concerns a method for screening for agents which are capable of inhibiting the cellular function of the nuclear receptor L66 comprising the steps of contacting one or more candidate agents with a polypeptide according to the invention, removing unbound agent(s) and detecting whether the agent(s) interact with the polypeptid of the nuclear receptor.

The invention also concerns method for inhibiting the cellular function of the nuclear receptor L66, comprising the steps of contacting a cell with a binding agent of a polypeptide previously identified as outlined herein whereby the cellular function of L66 is inhibited.

Such a binding agent may be an antibody., RNA. an anti-sense oligonucleotide. a ribozyme or one of substances shown below or identified in a respective assay as disclosed herein.

In still a further embodiment, the present invention concerns a method for identifying new nuclear receptor inhibitory or stimulatory substances, which may be termed as "candidate substances". It is contemplated that this screening technique proves useful in the general identification of compounds that serve the purpose of inhibiting or stimulating nuclear receptor activity.

In one embodiment of the invention the following substances are disclosed as potential interactors of the nuclear receptor according to the invention:

Steroids: dexamethasone-t-butylacetate, RU486, progesterone, 17-alpha-hydroxyprogesterone, 1,16-alpha dimethylpregnenolone, 17-alpha-hydroxypregnenonlone, pregnenonlone, 5beta-pregnane-3,20-dione, pregnenonlone-16-carbonitrile, 5beta-pregnane-3,20-dione, androstanol, corticosterone, dehydroepiandrosterone, dihydroxytestosterone, estradiol, cortisol, cortisone, dihydroxytestosterone.

Other substances: transnonachlor, chlordane, spironolactone, cyproterone acetate, rifampicin, nefipine, diethylstilbestrol, coumesterol, clotrimazole, lovastatin, phenoarbital, pthalic acid, nonylphenol, 1,4-bis(2-(3,5-dichloropyridyloxy1))benzene,

This also includes the use of heteromultimeric complexes of the nuclear receptor with other proteins, such as heterodimeric complexes with RXR, or any other binding partner.

Accordingly, in screening assays to identify pharmaceuticals agents which affect nuclear receptor activity, it is proposed that compounds isolated from natural sources, such as fungal extracts, plant extracts, bacterial extracts, higher eukaryotic cell extracts, or even extracts from animal sources, or marine, forest or soil samples, may be assayed for the presence of potentially useful pharmaceutical agents.

It will be understood that that the pharmaceutical agents to be screened could also be derived from chemical compositions or man-made compounds. The candidate substances can could also include monoclonal or polyclonal antibodies, peptides or proteins, such as those derived from recombinant DNA technology or by other means, including chemical peptide synthesis. The active compounds may include fragments or parts or derivatives of naturally-occurring compounds or may be only found as active combinations of known compounds which are otherwise inactive. We anticipate that such screens will in some cases lead to the isolation of agonists of nuclear receptors, in other cases to the isolation of antagonists. In other instances, substances will be identified that have mixed agonistic and antagonistic effects, or affect nuclear receptors in any other way.

CELL BASED ASSAYS

To identify a candidate substance capable of influencing L66 nuclear receptor activity, one first obtains a recombinant cell line. One designs the cell line in such a way that the activity of the nuclear receptor leads to the expression of a protein which has an easily detectable phenotype (a reporter), such as luciferase, fluorescent proteins such as green or red fluorescent protein, beta-galactosidase, alpha-galactosidase, beta-lactamase, chloramphenicol-acetyl-transferase, beta-glucuronidase, or any protein which can be detected by a secondary reagent such as an antibody.

Methods for detecting proteins using antibodies, such as ELISA assays, are well known to those skilled in the art.

Here, the amount of reporter protein present reflects the transcriptional activity of the nuclear receptor. This recombinant cell line is then screened for the effect of

substances on the expression of the reporters, thus measuring the ffect of these substances on the activity of the nuclear receptor. These substances can be derived from natural sources, such as fungal extracts, plant extracts, bacterial extracts, higher eukaryotic cell extracts, or even extracts from animal sources, or marine, forest or soil samples, may be assayed for the presence of potentially useful pharmaceutical agents. It will be understood that that the pharmaceutical agents to be screened may be derived from chemical compositions or man-made compounds.

The candidate substances can also include monoclonal or polyclonal antibodies, peptides or proteins, such as those derived from recombinant DNA technology or by other means, including chemical peptide synthesis. The active compounds may include fragments or parts or derivatives of naturally-occurring compounds or may be only found as active combinations of known compounds which are otherwise inactive.

In general the assay comprises, contacting a suitable cell containing a reporter under the control of the L66 nuclear receptor with a test compound, monitoring said host cell for the expression of the reporter gene, wherein expression of the reporter reflects the transcriptional activity of the nuclear receptor L66, and therefore reflects effects of the compound on the nuclear receptor.

In other embodiments of the invention assays are included where measuring the activity of a dimer of the nuclear receptor L66 and another protein, such as RXR takes place. Further included are assays aiming at the identification of compounds which specifically influence only the monomeric, homodimeric or homomultimeric form of the nuclear receptor, or influencing only multimeric forms of the nuclear receptor. Such assays include measuring the effect of a compound on the nuclear receptor in the absence of a binding partner, and measuring the effect of a compound on the nuclear receptor in the presence of a binding partner, such as RXR. One skilled in the art will find numerous more assays which are equally covered by the invention.

A cell line where the activity of a nuclear receptor determines the expression of a reporter can be obtained by creating a fusion gene driving the expression of a fusion

protein consisting of the ligand-binding domain of the L66 nuclear receptor fused to the DNA binding domain of a transcription factor with a known specificity for a given DNA sequence (the binding site). This DNA sequence (the binding site) can then be inserted in one or multiple copies before (5') to the promoter driving the expression of the reporter. Transcription factors useful for this approach include bacterial proteins, such as lexA, yeast proteins, such as Gal4, mammalian proteins such as NFkappaB or NFAT, the glucocorticoid receptor, the estrogen receptor, or plant proteins. The binding sites for these proteins can all be used in combination with the appropriate transcription factor to generate a useful reporter assay.

Another way to screen for inhibitors is to identify binding sites on DNA for the L66 nuclear receptor, and operatively link this binding site to a promoter operatively linked to a reporter gene. Included among others are binding sites for heterodimers of the L66 nuclear receptor with another protein, such as RXR.

Furthermore, transgenic animals described in the invention can be used to derive cell lines useful for cellular screening assays.

Cell lines useful for such an assay include many different kinds of cells, including prokaryotic, animal, fungal, plant and human cells. Yeast cells can be used in this assay, including *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* cells.

Another way to build cellular assays to measure the effect of compounds is the use of the yeast two hybrid system (see for example see, for example, U.S. Pat. No. 5,283,317; Zervos et al. (1993) Cell 72:223-232; Madura et al. (1993) J. Biol. Chem. 268:12046-12054; Bartel et al. (1993) Biotechniques 14:920-924; Iwabuchi et al. (1993) Oncogene 8:1693-1696; PCT Publication No. WO 94/10300, and U.S. Pat. No. 5,667,973), and or possible variants of the basic two hybrid system as discussed *e.g* in Vidal M, Legrain P, Nucleic Acids Res. 1999 Feb 15;27(4):919-29. Briefly, the two hybrid assay relies on reconstituting in vivo a functional transcriptional activator protein from two separate fusion proteins. In particular, the method makes use of chimeric genes which express hybrid proteins. To illustrate, a first hybrid gene comprises the coding sequence for a DNA-binding domain of a transcriptional activator fused in frame to the coding sequence for a TI polypeptide. The second

hybrid protein encodes a transcriptional activation domain fused in frame to a sample gene from a cDNA library. If the bait and sample hybrid proteins are able to interact, e.g., form a TI-dependent complex, they bring into close proximity the two domains of the transcriptional activator. This proximity is sufficient to cause transcription of a reporter gene which is operably linked to a transcriptional regulatory site responsive to the transcriptional activator, and expression of the reporter gene can be detected and used to score for the interaction of the TI and sample proteins.

In such assays, one primarily measures the effect of a compound on a given interaction involving the L66 nuclear receptor and a binding protein. In a preferred embodiment of the invention systems using other hosts such as prokaryotes as *E. coli*, or eukaryotic mammalian cells are described.

Two hybrid systems using hybrid protein fusions with other proteins than transcription factors, including enzymes such as beta-galactosidase or *dihydrofolate reductase* may also be applied. These assays are useful both to monitor the effect of a compound, including peptides, proteins or nucleic acids on an interaction of a nuclear receptor with a given binding partner, as well as to identify novel proteins or nucleic acids interacting with the nuclear receptor.

Monitoring the influence of compounds on cells may be applied not only in basic drug screening, but also in clinical trials. In such clinical trials, the expression of a panel of genes may be used as a "read out" of a particular drug's therapeutic effect.

CELL-FREE ASSAYS

Recombinant forms of the polypeptide according to SEQ ID NO. 3, 24 or 19 or as encoded by the nucleic acids according to the invention can be used in cell-free screening assays aiming at the isolation of compounds affecting the activity of nuclear receptors. In such an assay, the nuclear receptor polypeptide is brought into contact with a substance to test if the substance has an effect on the activity of the L66 receptor.

The detection of an interaction between an agent and a receptor may be accomplished through techniques well-known in the art. These techniques include but are not limited to centrifugation, chromatography, electrophoresis and spectroscopy. The use of isotopically labeled reagents in conjunction with these techniques or alone is also contemplated. Commonly used radioactive isotopes include ^{3}H, ^{14}C, . ^{22}Na, ^{32}P, ^{33}P, ^{35}S, ^{45}Ca, ^{60}Co, ^{125}I, and ^{131}I. Commonly used stable isotopes include ^{2}H, .^{13}C, ^{15}N, ^{18}O.

For example, if an agent binds to the receptor of the present invention, the binding may be detected by using radiolabeled agent or radiolabeled receptor. Briefly, if radiolabeled agent or radiolabeled receptor is utilized, the agent-receptor complex may be detected by liquid scintillation or by exposure to x-ray film or phosho-imaging devices.

One way to screen for substances affecting nuclear receptor activity is to measure the effect of the binding of nuclear receptors to ligands, such as cofactors, activators, repressors, DNA, RNA, proteins, antibodies, peptides or other substances, including chemical compounds known to affect receptor activity. Assays measuring the binding of a protein to a ligand are well known in the art, such as ELISA assays, FRET assays, bandshift assays, plasmon-resonance based assays, scintilllation proximity assays, fluorescence polarization assays.

In one example, a mixture containing the L66 polypeptide, effector and candidate substance is allowed to incubate. The unbound effector is separable from any effector/receptor complex so formed. One then simply measures the amount of each (*e.g.*, versus a control to which no candidate substance has been added). This measurement may be made at various time points where velocity data is desired. From this, one determines the ability of the candidate substance to alter or modify the function of the receptor.

Numerous techniques are known for separating the effector from effector/receptor complex, and all such methods are intended to fall within the scope of the invention. This includes the use of thin layer chromatographic methods (TLC), HPLC, spectrophotometric, gas chromatographic/mass spectrophotometric or NMR

analyses. Another method of separation is to immobilize one of the binding partners on a solid support, and to wash away any unbound material. It is contemplated that any such technique may be employed so long as it is capable of differentiating between the effector and complex, and may be used to determine enzymatic function such as by identifying or quantifying the substrate and product.

A screening assay provides a L66 receptor under conditions suitable for the binding of an agent to the L66 receptor. These conditions include but are not limited to pH, temperature, tonicity, the presence of relevant cofactors, and relevant modifications to the polypeptide such as glycosylation or lipidation. It is contemplated that the receptor can be expressed and utilized in a prokaryotic or eukaryotic cell. The host cell expressing the L66 receptor can be used whole or the receptor can be isolated from the host cell. The L66 receptor can be membrane bound in the membrane of the host cell or it can be free in the cytosol of the host cell. The host cell can also be fractionated into sub-cellular fractions where the receptor can be found. For example, cells expressing the receptor can be fractionated into the nuclei, the endoplasmic reticulum, vesicles, or the membrane surfaces of the cell.

pH is preferably from about a value of 6.0 to a value of about 8.0, more preferably from about a value of about 6.8 to a value of about 7.8, and most preferably, about 7.4. In a preferred embodiment, temperature is from about 20°C degrees to about 50°C degrees more preferably, from about 30°C degrees to about 40°C degrees and even more preferably about 37°C degrees. Osmolality is preferably from about 5 milliosmols per liter (mosm/L) to about 400 mosm/l, and more preferably, from about 200 milliosmols per liter to about 400 mosm/l and, even more preferably from about 290 mosm/L to about 310 mosm/L. The presence of cofactors can be required for the proper functioning of the L66 receptor. Typical cofactors include sodium, potassium, calcium, magnesium, and chloride. In addition, small, non-peptide molecules, known as prosthetic groups may also be required. Other biological conditions needed for receptor function are well-known in the art.

It is well-known in the art that proteins can be reconstituted in artificial membranes, vesicles or liposomes. (Danboldt et al.,1990). The present invention contemplates

that the receptor can be incorporated into artificial membranes, vesicles or liposomes. The reconstituted receptor can be utilized in screening assays.

It is further contemplated that a receptor of the present invention can be coupled to a solid support, *e.g.*, to agarose beads, polyacrylamide beads, polyacrylic, sepharose beads or other solid matrices capable of being coupled to polypeptides. Well-known coupling agents include cyanogen bromide (CNBr), carbonyldiimidazole, tosyl chloride, diaminopimelimidate, and glutaraldehyde.

In a typical screening assay for identifying candidate substances, one employs the same recombinant expression host as the starting source for obtaining the receptor polypeptide, generally prepared in the form of a crude homogenate. Recombinant cells expressing the receptor are washed and homogenized to prepare a crude polypeptide homogenate in a desirable buffer such as disclosed herein. In a typical assay, an amount of polypeptide from the cell homogenate, is placed into a small volume of an appropriate assay buffer at an appropriate pH. Candidate substances, such as agonists and antagonists, are added to the admixture in convenient concentrations and the interaction between the candidate substance and the receptor polypeptide is monitored.

Where one uses an appropriate known substrate for the L66 receptor, one can, in the foregoing manner, obtain a baseline activity for the recombinantly produced L66 receptor. Then, to test for inhibitors or modifiers of the receptor function, one can incorporate into the admixture a candidate substance whose effect on the L66 receptor is unknown. By comparing reactions which are carried out in the presence or absence of the candidate substance, one can then obtain information regarding the effect of the candidate substance on the normal function of the receptor.

Accordingly, this aspect of the present invention will provide those of skill in the art with methodology that allows for the identification of candidate substances having the ability to modify the action of nuclear receptor polypeptides in one or more manners.

Additionally, screening assays for the testing of candidate substances are designed to allow the determination of structure-activity relationships of agonists or antagonists

with the receptors, e.g., comparisons of binding between naturally-occurring hormones or other substances capable of interacting with or otherwise modulating the receptor; or comparison of the activity caused by the binding of such molecules to the receptor.

In certain aspects, the polypeptides of the invention are crystallized in order to carry out x-ray crystallographic studies as a means of evaluating interactions with candidate substances or other molecules with the nuclear receptor polypeptide. For instance, the purified recombinant polypeptides of the invention, when crystallized in a suitable form, are amenable to detection of intra-molecular interactions by x-ray crystallography. In another aspect, the structure of the polypeptides can be determined using nuclear magnetic resonance.

PHARMACEUTICAL COMPOSITION:

This invention provides a pharmaceutical composition comprising an effective amount of a agonist or antagonist drug identified by the method described herein and a pharmaceutically acceptable carrier. Such drugs and carrier can be administered by various routes, for example oral, subcutaneous, intramuscular, intravenous or intracerebral. The preferred route of administration would be oral at daily doses of about 0.01 -100 mg/kg.

This invention provides a method of treating metabolic disorders, immunological indications, hormonal dysfunctions, neurosystemic diseases wherein the abnormality is improved by reducing the activity of L66 receptor or blocking the binding of ligands to a L66 receptor, which method comprises administering an effective amount of the antagonist-containing pharmaceutical composition described above to suppress the subject's appetite. Similarly, the invention also provides methods for treating diseases and conditions resulting from metabolic disorders, immunological indications, hormonal dysfunctions, neurosystemic diseases, which method comprises administering an effective amount of an agonist-containing pharmaceutical composition described above.

TRANSFORMATION OF CELLS AND DRUG SCREENING :

The recombinant expression constructs of the present invention are useful in molecular biology to transform cells which do not ordinarily express L66 to thereafter express this receptor.

Such cells are useful as intermediates for making cellular preparations useful for receptor binding assays, which are in turn useful for drug screening. Drugs identified from such receptor assays can be used for the treatment of metabolic disorders, immunological indications, hormonal dysfunctions, and/or neurosystemic diseases.

The recombinant expression constructs of the present invention are also useful in gene therapy. Cloned genes of the present invention, or fragments thereof, may also be used in gene therapy carried out by homologous recombination or site-directed mutagenesis. See generally Thomas & Capecchi, Cell 51, 503-512 (1987); Bertling, Bioscience Reports 7, 107-112 (1987); Smithies et al., Nature 317, 230-234 (1985).

Oligonucleotides of the present invention are useful as diagnostic tools for probing L66 expression in tissues. For example, tissues are probed *in situ* with oligonucleotide probes carrying detectable groups by conventional autoradiographic techniques, as explained in greater detail in the Examples below, to investigate native expression of this receptor or pathological conditions relating thereto. Further, chromosomes can be probed to investigate the presence or absence of the L66, and potential pathological conditions related thereto, as also illustrated by the Examples below. Probes according to the invention should generally be at least about 15 nucleotides in length to prevent binding to random sequences, but, under the appropriate circumstances may be smaller (see above for details on hybridization).

ANTIBODIES AGAINST THE L66 NUCLEAR RECEPTOR PROTEIN OR POLYPEPTIDE

Another aspect of the invention includes an antibody specifically reactive with the protein or any part of the protein according to the invention (SEQ ID NO. 3, 24 or 19)

and or a polypeptide encoded by the nucleotide sequence of the nuclear receptor L66 (see also figures). (The term „antibody" refers to intact molecules as well as fragments thereof, such as Fa, F(ab).sub.2, and Fv, which are capable of binding the epitopic determinant.) By using immunogens derived from the polypeptide according to the invention (SEQ ID NO. 3, 24, 19) and/or encoded by the nucleic acids according to the invention, anti-protein/anti-peptide antisera or monoclonal antibodies can be made by standard protocols (E. Howell & D. Lane. Antibodies: A Laboratory Manual. *Cold Spring Harbor Laboratory* (1988)).

A polyclonal antibody is prepared by immunizing a mammal, such as a mouse, a hamster or rabbit with an immunogenic form of the polypeptide, i.e. the human L66 polypeptide of the present invention, and collecting antisera from that immunized animal. Because of the relatively large blood volume of rabbits, a rabbit is a preferred choice for production of polyclonal antibodies.

As an immunizing antigen, fusion proteins, intact polypeptides or fragments containing small peptides of interest can be used. They can be derived by expression from a cDNA transfected in a host cell with subsequent recovering of the protein/peptide or peptides can be synthesized chemically (*e.g.* oligopeptides with 10-15 residues in length). Important tools for monitoring the function of the gene according to the present invention, *i.e.* encoded by a sequence according to SEQ ID NO. 1, 4, 24 or 17 (or portions thereof or splice variants thereof) are antibodies against various domains of the protein according to the invention. Various Oligopeptides from the N- and C-terminal sequences and the DBD/hinge region of the protein can be used as antigens.

A given polypeptide or polynucleotide may vary in its immunogenicity. It is often necessary to couple the immunogen (e.g. the polypeptide) with a carrier. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin (BSA) and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal in the presence of an adjuvant, a non-specific stimulator of the immune response in order to enhance immunogenicity. The production of polyclonal antibodies is monitored by detection of antibody titers in plasma or serum at various time points following immunization. Standard ELISA or other immunoassays can be

used with the immunogen as antigen to assess the lev ls of antibodies. When a desired level of immunogenicity is obtained, the immunized animal may be bled and the serum isolated, stored and purified.

To produce monoclonal antibodies, antibody-producing cells (e.g. spleen cells) from an immunized animal (preferably mouse or rat) are fused by standard somatic cell fusion procedures with immortalizing cells such as myeloma cells to yield hybridoma cells. Where the immunized animal is a mouse, a preferred myeloma cell is the murine NS-1 myeloma cell. Such techniques are well known in the art, and include, for example, the hybridoma technique (originally developed by Kohler & Milstein. *Nature* 256: 495-497 (1975)), the human B cell hybridoma technique (Kozbar *et al. Immunology Today* 4:72 (1983)), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole *et al.* Monoclonal Antibodies and Cancer Therapy. *Alan R. Liss, Inc.* pp. 77-96 (1985)).

The fused spleen/myeloma cells are cultured in a selective medium to select fused spleen/myeloma cells from the parental cells. Fused cells are separated from the mixture of non-fused parental cells, for example, by the addition of agents that block the *de novo* synthesis of nucleotides in the tissue culture media. This culturing provides a population of hybridomas from which specific hybridomas are selected. Typically, selection of hybridomas is performed by culturing the cells by single-clone dilution in microtiter plates, followed by testing the individual clonal supernatants for reactivity with an antigen-polypeptide. The selected clones may then be propagated indefinitely to provide the monoclonal antibody in convenient quantity.

The creation of antibodies which specifically bind the polypeptide according to the invention (SEQ ID NO. 3, 24 or 19) and/or encoded by the nucleotide sequence of the nuclear receptor L66 or its complement (SEQ ID NO. 1, 4, 17 or 22) provides an important utility in immunolocalization studies, and may play an important role in the diagnosis and treatment of receptor disorders. The antibodies may be employed to identify tissues, organs, and cells which express or the nuclear receptor L66. Antibodies can be used diagnostically in immuno-precipitation and immuno-blotting to detect and evaluate nuclear receptor L66 protein levels in tissue or from cells in bodily fluid as part of a clinical testing procedure.

Monoclonal antibodies provided by the present invention are also produced by recombinant genetic methods well known to those of skill in the art, and the present invention encompasses antibodies made by such methods that are immunologically reactive with an epitope of a mammalian nuclear L66 receptor protein or peptide.

The present invention encompasses fragments of the antibody that are immunologically reactive with an epitope of a mammalian nuclear L66 receptor protein or peptide. Such fragments are produced by any number of methods, including but not limited to proteolytic cleavage, chemical synthesis or preparation of such fragments by means of genetic engineering technology. The present invention also encompasses single-chain antibodies that are immunologically reactive with an epitope of a mammalian nuclear L66 receptor protein or peptide made by methods known to those of skilled in the art.

CHIMERIC ANTIBODIES AND OTHER TYPES OF ANTIBODIES:

The invention also includes chimeric antibodies, comprised of light chain and heavy chain peptides immunologically reactive to an epitope that is a mammalian nuclear L66 receptor protein or peptide. The chimeric antibodies embodied in the present invention include those that are derived from naturally occurring antibodies as well as chimeric antibodies made by means of genetic engineering technology well known to those of skill in the art.

Also included are methods for the generation of antibodies against L66 which rely on the use of phage display systems and related systems, such as described in Hoogenboom HR, de Bruine AP, Hufton SE, Hoet RM, Arends JW, Roovers RC, Immunotechnology 1998 Jun;4(1):1-20, and references therein.

EPITOPES OF THE L66 NUCLEAR RECEPTOR

The present invention also encompasses an epitope of a mammalian nuclear L66 receptor protein or peptide that is comprised of sequences and/or a conformation of

sequences present in the mammalian nuclear L66 receptor protein or peptide molecule. This epitope may be naturally occurring, or may be the result of proteolytic cleavage of the mammalian nuclear L66 receptor protein or peptide molecule and isolation of an epitope-containing peptide or may be obtained by synthesis of an epitope-containing peptide using method of genetic engineering technology and synthesized by genetically engineered prokaryotic or eukaryotic cells.

ANTISENSE OLIGONUCLEOTIDES AGAINST L66

Antisense oligonucleotides are short single stranded DNA or RNA molecules which may be used to block the availability of the L66 receptor messenger. Synthetic derivatives of ribonucleotides or deoxyribonucleotides and/or PNAs (see above) are equally possible.

The sequence of an antisense oligonucleotide is at least partially complementary to the sequence (or the gene) of interest. The complementarity of the sequence is in any case high enough to enable the antisense oligonucleotide to bind to the nucleic acid according to the invention or parts thereof. Many examples exist in which the binding of oligonucleotides to the target sequence interfere with the biological function of the targeted sequence (Brysch W, Schlingensiepen KH, Design and application of antisense oligonucleotides in cell culture, in vivo, and as therapeutic agents, Cell Mol Neurobiol 1994 Oct;14(5):557-68; Wagner RW, Gene inhibition using antisense oligodeoxynucleotides, Nature 1994 Nov 24;372(6504):333-5 or Brysch W, Magal E, Louis JC, Kunst M, Klinger I, Schlingensiepen R, Schlingensiepen KH Inhibition of p185c-erbB-2 proto-oncogene expression by antisense oligodeoxynucleotides down-regulates p185-associated tyrosine-kinase activity and strongly inhibits mammary tumor-cell proliferation, Cancer Gene Ther 1994 Jun;1(2):99-105 or Monia BP, Johnston JF, Ecker DJ, Zounes MA, Lima WF, Freier SM Selective inhibition of mutant Ha-ras mRNA expression by antisense oligonucleotides, J Biol Chem 1992 Oct 5;267(28):19954-62 or Bertram J, Palfner K, Killian M, Brysch W, Schlingensiepen KH, Hiddemann W, Kneba M, Reversal of multiple drug resistance in vitro by phosphorothioate oligonucleotides and ribozymes, Anticancer Drugs 1995 Feb;6(1):124-34)

This interference occurs in most instances at the level of translation, *i.e.* through the inhibition of th translational machinery by oligonucleotides that bind to mRNA, however, two other mechanisms of interference with a given gene's function by oligonucleotides can also be envisioned, (i) the functional interference with the transcription of a gene through formation of a triple helix at the level of genomic DNA and the interference of oligonucleotides with the function of RNA molecules that are executing at least part of their biological function in the untranslated form (Kochetkova M, Shannon MF, Triplex-forming oligonucleotides and their use in the analysis of gene transcription. Methods Mol Biol 2000;130:189-201 Rainer B. Lanz1, Neil J. McKenna1, Sergio A. Onate1, Urs Albrecht2, Jiemin Wong1, Sophia Y. Tsai1, Ming-Jer Tsai1, and Bert W. O'Malley A Steroid Receptor Coactivator, SRA, Functions as an RNA and Is Present in an SRC-1 Complex Cell, Vol. 97, 17–27, April, 1999).

Antisense oligonucleotides can be conjugated to different other molecules in order to deliver them to the cell or tissue expressing L66. For instance the antisense oligonucleotide can be conjugated to a carrier protein (*e.g.* ferritin) in order to direct the oligonucleotide towards the desired target tissue, *i.e.* in case of ferritin predominantly to the liver.

Antisense expression constructs are expression vector systems that allow the expression – either inducible or uninducible - of a complementary sequence to the L66 sequences according to the invention. The potential possibility of such an approach has been demonstrated in many different model systems (von Ruden T, Gilboa E, Inhibition of human T-cell leukemia virus type I replication in primary human T cells that express antisense RNA, J Virol 1989 Feb;63(2):677-82; Nemir M, Bhattacharyya D, Li X, Singh K, Mukherjee AB, Mukherjee BB, Targeted inhibition of osteopontin expression in the mammary gland causes abnormal morphogenesis and lactation deficiency, J Biol Chem 2000 Jan 14;275(2):969-76; Ma L, Gauville C, Berthois Y, Millot G, Johnson GR, Calvo F.Antisense expression for amphiregulin suppresses tumorigenicity of a transformed human breast epithelial cell line, Oncogene 1999 Nov 11;18(47):6513-20; Refolo LM, Eckman C, Prada CM, Yager D, Sambamurti K, Mehta N, Hardy J, Younkin SG, Antisense-induced reduction of presenilin 1 expression selectively increases the production of amyloid beta42 in

transfected cells, J Neurochem 1999 Dec;73(6):2383-8; Buckley NJ, Abogadie FC, Brown DA, Dayrell M, Caulfield MP, Delmas P, Haley JE, Use of antisense expression plasmids to attenuate G-protein expression in primary neurons, Methods Enzymol 2000;314:136-48).

According to the invention an antisense expression construct can be constructed with virtually any expression vector capable of fulfilling at least the basic requirements known to those skilled in the art.

In one embodiment of the invention retroviral expression systems or tissue specific gene expression systems are preferred.

Current standard technologies for delivering antisense constructs are performed through a conjugation of constructs with liposomes and related, complex-forming compounds, which are transferred via electroporation techniques or via particle-mediated "gene gun" technologies into the cell. Other techniques may be envisioned by one skilled in the art.

Microinjection still plays a major role in most gene transfer techniques for the generation of germ-line mutants expressing foreign DNA (including antisense RNA constructs) and is preferred embodiment of the present invention.

RIBOZYMES DIRECTED AGAINST L66

Ribozymes are either RNA molecules (Gibson SA, Pellenz C, Hutchison RE, Davey FR, Shillitoe EJ, Induction of apoptosis in oral cancer cells by an anti-bcl-2 ribozyme delivered by an adenovirus vector, Clin Cancer Res 2000 Jan;6(1):213-22; Folini M, Colella G, Villa R, Lualdi S, Daidone MG, Zaffaroni N, Inhibition of Telomerase Activity by a Hammerhead Ribozyme Targeting the RNA Component of Telomerase in Human Melanoma Cells, J Invest Dermatol 2000 Feb;114(2):259-267; Halatsch ME, Schmidt U, Botefur IC, Holland JF, Ohnuma T, Marked inhibition of glioblastoma target cell tumorigenicity in vitro by retrovirus-mediated transfer of a hairpin ribozyme against deletion-mutant epidermal growth factor receptor messenger RNA, J Neurosurg 2000 Feb;92(2):297-305; Ohmichi T, Kool ET, The virtues of self-binding:

high sequence specificity for RNA cleavage by self-processed hammerhead ribozymes, Nucleic Acids Res 2000 Feb 1;28(3):776-783) or DNA molecules (Li J, Zheng W, Kwon AH, Lu Y, In vitro selection and characterization of a highly efficient Zn(II)-dependent RNA-cleaving deoxyribozyme; Nucleic Acids Res 2000 Jan 15;28(2):481-488) that have catalytic activity. The catalytic activity located in one part of the RNA (or DNA) molecule can be "targeted" to a specific sequence of interest by fusing the enzymatically active RNA molecule sequence with a short stretch of RNA (or DNA) sequence that is complementary to the L66 transcript. Such a construct will, when introduced into a cell either physically or via gene transfer of a ribozyme expression construct find the L66 sequence (our sequence of interest) and bind via its sequence-specific part to said sequence. The catalytic activity attached to the construct, usually associated with a special nucleic acid structure (people distinguish so called "hammerhead" structures and "hairpin" structures), will then cleave the targeted RNA. The targeted mRNA will be destroyed and cannot be translated efficiently, thus the protein encoded by the mRNA derived from L66 will not be expressed or at least will be expressed at significantly reduced amounts.

In a preferred embodiment the invention covers inducible ribozyme constructs (Koizumi M, Soukup GA, Kerr JN, Breaker RR, Allosteric selection of ribozymes that respond to the second messengers cGMP and cAMP, Nat Struct Biol 1999 Nov;6(11):1062-1071).

In a further preferred embodiment the invention concerns the use of "bivalent" ribozymes (multimers of catalytically active nucleic acids) as described in (Hamada M, Kuwabara T, Warashina M, Nakayama A, Taira K, Specificity of novel allosterically trans- and cis-activated connected maxizymes that are designed to suppress BCR-ABL expression FEBS Lett 1999 Nov 12;461(1-2):77-85).

TRANSGENIC ANIMALS CARRYING THE L66 NUCLEAR RECEPTOR

Also provided by the present invention are non-human transgenic animals grown from germ cells transformed with the L66 nucleic acid sequence according to the invention and that express the L66 receptor according to the invention and offspring and descendants thereof. Also provided are transgenic non-human mammals

comprising a homologous recombination knockout of the native L66 receptor, as well as transgenic non-human mammals grown from germ cells transformed with nucleic acid antisense to the L66 nucleic acid of the invention and offspring and descendants thereof. Further included as part of the present invention are transgenic animals which the native L66 receptor has been replaced with the human homolog. Of course, offspring and descendants of all of the foregoing transgenic animals are also encompassed by the invention.

Transgenic animals according to the invention can be made using well known techniques with the nucleic acids disclosed herein. E.g., Leder et al., U.S. Patent Nos.4,736,866 and 5,175,383; Hogan et al., Manipulating the Mouse Embryo, A Laboratory Manual (Cold Spring Harbor Laboratory (1986)); Capecchi, Science 244, 1288 (1989); Zimmer and Gruss, Nature 338, 150 (1989); Kuhn et al., Science 269, 1427 (1995); Katsuki et al., Science 241, 593 (1988); Hasty et al., Nature 350, 243 (1991); Stacey et al., Mol. Cell Biol. 14, 1009 (1994); Hanks et al., Science 269, 679 (1995); and Marx, Science 269, 636 (1995). Such transgenic animals are useful for screening for and determining the physiological effects of L66 receptor agonists and antagonist.

Consequently, such transgenic animals are useful for developing drugs to regulate physiological activities in which L66 participates.

The following examples are provided for illustrative purposes only and are not intended, nor should they be construed, as limiting the invention in any manner.

MODELLING OF THE STRUCTURE OF L66

The novel nuclear receptor sequences disclosed herein may be used for various in silico, i.e. computer analyses. Such analyses may be for example nuclear receptor specific sequence alignments which permit the identification of domains and even new receptors. The novel domain sequences disclosed herein may be used in order to create domain specific hidden markov models (hmms) or simply as search sequences.

In a preferred embodiment this similarity search tool is the BLAST algorithm.(Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman (1990). Basic local alignment search tool. J. Mol. Biol. 215:403-10 and the sequence used is one of those disclosed herein.

Another search tool that may be used is FASTA (W. R. Pearson and D. J. Lipman (1988), "Improved Tools for Biological Sequence Analysis", PNAS 85:2444- 2448, and W. R. Pearson (1990) "Rapid and Sensitive Sequence Comparison with FASTP and FASTA" Methods in Enzymology 183:63- 98).

The invention is not limited to one particular type of search tool. In one embodiment of the invention search tools are used that do not search by sequence similarity but by applying sequence profiles such as a profile generated when applying the Profile Hidden Markov Model.

Profile Hidden Markov Models also called „Hidden Markov Models“, here abbreviated as HMM, are statistical models representing the consensus of the primary structure of a sequence family. The profiles use scores specific of the position of amino acids (or nucleotides) and position specific scores for the opening or the expansion of an insertion or deletion. Methods for the creation of profiles, starting from multiple alignments, have been introduced by Taylor (1986), Gribskov et al. (1987), Barton (1990) and Heinikoff (1996).

HMMs provide an utterly probabilistic description of profiles, *i.e.* Bayes' theory rules the positioning of all probability (evaluation) parameters (compare Krogh et al. 1994, Eddy 1996 und Eddy 1998). The central idea behind this is that a HMM is a finite model describing the probability distribution of an infinite number of possible sequences. The HMM consists of a number of states corresponding with the columns of a multiple alignment as it is usually depicted. Each state emits symbols (remainders) corresponding with the probability of the symbol emission (specific of the respective state), and the states are linked with each other by probabilities of the changing of states. Starting from one specific state, a succession of states is generated by changing from one state to the other in accordance with the probability

of the changing of states, until a final state has been reached. Each state then emits symbols according to the probability distribution of emissions specific of this state, creating an observable sequence of symbols.

The attribute „hidden" has been derived from the fact that the underlying sequence of states cannot be observed. Only the sequence of symbols is visible. An assessment of the probabilities of changing of states and of emissions (the training of the model) is achieved by dynamic programming algorithms implemented in the HMMER package.

The sequences according to the invention may be aligned with other nuclear receptor sequences in order to create a multiple sequence alignment which is used as the basis for the creation of a HMM.

If an existing HMM and a sequence are given, the probability that the HMM could generate the sequence in question, can be calculated. The HMMER package provides a numerical quantity (the Score) in proportion to this probability, *i.e.* the information content of the sequence indicated as bits, measured according to the HMM.

See also Barton, G.J. (1990): Protein multiple alignment and flexible pattern matching, Methods Ezymol. 183: 403-427, Eddy, S.R. (1996): Hidden markov models. Curr. Opin. Strct. Biol. 6: 361-365, Eddy, S.R. (1998): Profile hidden markov models.Bioinformatics. 14: 755-763, Gribskov, M. McLachlan, A.D. und Eisenberg D. (1987): Profile analysis: Detection of distantly related proteins. Proc. Natl. Acad. Sci. USA. 84: 4355-5358, Heinikoff, S. (1996): Scores for sequence searches and alignment, Curr. Opin. Strct. Biol. 6: 353-360, Krogh, A., Brown, M., Mian, I.S., Sjolander, K. und Haussler, D. (1994): Hidden markov models in computational biology: Applications to protein modelling. J. Mol. Biol. 235: 1501-1531, Taylor, W.R. (1986): Identification of protein sequence homology by consensus template alignment. J. Mol. Biol. 188: 233-258.

In general the sequence are selected such that a query using a search sequence returns a result consisting of sequences which are at least to a certain degree similar to the query sequence.

In one embodiment of the invention amino acid sequences of the present invention are used to model the three-dimensional structure of the protein. Initially, this involves the comparison of the protein sequence with the sequence of related proteins where the structure is known, such as the human PPARγ ligand-binding domain (Nolte RT, Wisely GB, Westin S, Cobb JE, Lambert MH, Kurokawa R, Rosenfeld MG, Willson TM, Glass CK, Milburn MV,Nature 1998 Sep 10;395(6698):137-43). The three-dimensionale structure can then be modelled using computer programs. From the three-dimensional structure, binding sites of potential inhibitors or activators can be predicted. It can further be predicted which kinds of molecule might bind there. The predicted substances can then be screened to test their effect on nuclear receptor activity.

EXAMPLES

EXAMPLE 1: CLONING AND EXPRESSION OF THE GENE ACCORDING TO THE INVENTION

Construction of suitable vectors containing the desired coding and control sequences employs standard ligation and restriction techniques that are well understood in the art. Isolated plasmids, DNA sequences, were synthesized oligonucleotides were cleaved, tailored, and religated in the form desired.

Site-specific DNA cleavage was performed by treatment with the suitable restriction enzyme (or enzymes) under conditions that are generally understood in the art, and the particulars of which are specified by the manufacturer of these commercially available restriction nzymes.

See, e.g., New England Biolabs, Product Catalog. In general, about 1 μg of plasmid and/or DNA sequence was cleaved by one unit of enzyme in about 20 μl of buffer

solution. Often excess of restriction enzyme was used to ensure complete digestion of the DNA substrate. Incubation times of about one hour to two hours at about 37°C are workable, although variations are tolerable.

After each incubation, protein was removed by extraction with phenol/chloroform, and may be followed by ether extraction. The nucleic acid was recovered from aqueous fractions by precipitation with ethanol. If desired, size separation of the cleaved fragments was performed by polyacrylamide gel or agarose gel electrophoresis using standard techniques. A general description of size separations is found in Methods in Enzymology 65, 499-560 (1980).

Transformed host cells are cells which have been transformed or transfected with recombinant expression constructs made using recombinant DNA techniques and comprising mammalian nuclear receptor L66 encoding sequences. Preferred host cells for transient transfection are COS-7 cells. Transformed host cells may ordinarily express nuclear receptor L66, but host cells transformed for purposes of cloning or amplifying nucleic acid hybridization probe DNA need not express the nuclear L66 receptor. When expressed, the mammalian nuclear L66 receptor protein was typically located in the host cell membrane.

Cultures of cells derived from multicellular organisms are desirable hosts for recombinant nuclear receptor L66 protein synthesis. In principal, any higher eukaryotic cell culture is workable, whether from vertebrate or invertebrate culture. However, mammalian cells are preferred. Propagation of such cells in cell culture has become a routine procedure. See Tissue Culture (Academic Press, Kruse & Patterson, Eds., 1973). Examples of useful host cell lines are bacteria cells, insect cells, yeast cells, human 293 cells, VERO and HeLa cells, LMTK- cells, and WI138, BHK, COS-7, CV, and MDCK cell lines. Human 293 cells are preferred.

EXAMPLE 2: NUCLEAR RECEPTOR L66 TISSUE LOCALIZATION:

A multiple tissue northern blot (Clontech, Palo Alto) was hybridized to a labeled probe. The blot contained about 03 to 3 µg of poly A RNA derived from various tissues. Hybridization was carried out in a hybridization solution such as one

containing SSC (see Maniatis et al, ibid) at an optimized temperature between 50°c and 70°C, preferably 65°C. The filter was washed and a film xposed for signal detection (see also: Maniatis et al., *Molecular Cloning: A laboratory Manual*, Cold Spring Harbor Laboratory Press, N.Y.(1989)).

EXAMPLE 3: NUCLEAR RECEPTOR L66 cDNA ISOLATION FROM HUMAN AND OTHER ORGANISMS:

A cloning strategy was used to clone the L66 receptor cDNA from specific cDNA libraries (Clontech, Palo Alto) or alternatively, RNA was obtained from various tissues and used to prepare cDNA expression libraries by using for example an Invitrogen kit. (Invitrogen Corporation, San Diego). For the isolation of the L66 cDNA clone the chosen library was screened under stringent condition (see definitions above) by using an L66 specific probe. The cDNA insert of positive clones was subsequently sequenced and cloned in a suitable expression vector.

Additionally, full length receptor L66 clones from human and mouse (Mus musculus) was obtained by using RACE PCR technology. In brief, suitable cDNA libraries were constructed or purchased. Following reverse transcription, the first strand cDNA was used directly in RACE PCR reactions using a RACE cDNA amplification kit according to the manufactures protocol (Clontech, Palo Alto). Amplified fragments were purified, cloned and subsequently used for sequence analysis.

To obtain information about the genomic organization (see also figures) of the receptor gene, genomic libraries were screened (commercially available clone libraries were used) with a receptor specific probe under stringent conditions. Positive clones were isolated and the complete DNA sequence of the putative receptor region was determined by sequence analysis (Maniatis et al., *Molecular Cloning: A laboratory Manual*, Cold Spring Harbor Laboratory Press, N.Y.(1989)) (human genomic clone deposited at "Deutsche sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ, International Depositary Authority under the Budapest Treaty in Germany) under number DSM 14483).

EXAMPLE 4: NUCLEAR RECEPTOR L66 LIGAND BINDING ASSAY:

Saturation ligand binding analysis and ligand competition studies are carried out by using an over expressed L66 receptor-protein which is incubated with a labeled potential ligand at different concentrations. The L66 receptor may be immobilized on suitable surfaces. Various competitors are added in the presence of a labeled ligand. Bound and unbound ligand is separated be using e.g. gel filtration or charcoal based methods or simply removed by a washing step if the receptor protein L66 is immobilized. Receptor-bound ligand is detected by scintilation counting.

EXAMPLE 5: NUCLEAR RECEPTOR L66 EXPRESSION ASSAY:

Relative quantification was performed in multiplex PCR reactions using the ABI PRISM® 7700 Sequence Detection System. Total RNA was used as template and was either ordered from Clontech or Ambion in the case of human normal tissue or was isolated from cell lines that were ordered from the DSMZ (German Collection of Microorganisms and Cell Cultures).

Relative quantification was achieved by normalising the results for the presence of 18S rRNA in the samples and subsequently by standardising the corresponding expression data to the expression levels detected in testis (testis = 1). A standard curve was generated by diluting total RNA from testis tissue 1:10 fold, starting from 100 ng input amount and performing 7 dilution steps. For the detection of target mRNA in other samples the input amount was always 100 ng total RNA, samples were measured in triplicates.

For the detection of 18S rRNA the endogenous control pre-developed assay reagent "Ribosomal RNA control (18S rRNA)" from Applied biosystems was used. For the detection of the target sequence, PCR primers and the probe were designed according to Applied biosystem's specifications using its Primer Express® software. PCR primers were ordered from Interactiva (forward primer (SEQ ID NO. 14): CGT GGG CTA ATG AAT TTT ACC AAG; reverse primer (SEQ ID NO. 15): GGC CCC

ATG GAG AAA TAT CAC T). The probe (SEQ ID NO. 16) CCA ATG AGG ATC AAA CTG CAC TAC AGA AGG G was ordered from Applied Biosystems and was labelled with FAM at the 5'end and contained a quencher (TAMRA) at the 3' end.

EXAMPLE 6: FORMATION OF L66-RXR OR OTHER PROTEIN COMPLEXES

In order to explore the functional properties of L66, the DNA binding properties of L66 are analyzed. It has previously been shown that RXR is a common heterodimeric partner of various members of the nuclear receptor superfamily required for high affinity DNA binding (Hallenbeck et al., PNAS 89: 5572-5576 (1989)), Kliewer et al., EMBO J. 11: 1419-1435 (1992)). It has also been shown that DNA and ligand binding activities of the Dros. melanogaster ecdyson receptor (EcR) require heterodimer formation with RXR or USP (the homologue of RXR) (O'Malley in Endocrinology 125: 1119-1120 (1989)).

Consequently, it is of interest whether L66 can interact with RXR, or with other members of the family.

A two hybrid system is used. CV-1 cells are transiently transfected with cytomegalovirus promoter driven expression vectors containing the yeast GAL4 DNA binding domain (DBD) alone, GAL4 linked to L66 LBD (LBD; i.e. GAL4-L66) and the 78 amino acid Herpes virus VP16 transactivation domain (VP) linked to the amon terminal end of the LBDs for human (or mouse) RXRα (VP-RXR), mouse PPARgamma (VP-PPAR) VDR (VP-VDR) and others.

All cells are cotransfected with a luciferase reporter construct containing about 4 copies of the yeast GAL4 upstream activating sequence and a β-galactosidase expression vector as internal control.

CV-1 cells are grown in DMEM supplemented with 10% AG1-X8 resin-charcoal stripped calf bovine serum, 50 U/ml penicillin G and 50 µg/ml streptomycin sulfate (DMEM-CBS) at 37°C in 5% CO_2. One day prior to transfection, cells are plated to 50-80% confluence using a phenol-red free DMEM with 10% resin charcoal stripped

fetal bovine serum (DMEM-FBS). Cells are transfected (with reporter construct (300 ng/10^5 cells), cytomegalovirus driv n receptor (100 ng/10^5 cells) and β-galactosidase expression vectors (500 ng/10^5 cells) by lipofection using N-{2-(2,3)-dioleoyloxy)propyl-N,N,N-trimethyl ammonium methyl sulfate} according to the manufacturer's instructions (DOTAP, Boehringer Mannheim).

After 2 hours the liposomes are removed and cells treated for 40 hours with phenol-red free DMEM-FBS containing farnesol as the ligand. Cells are harvested and assayed for luciferase and β-galactosidase activity.

All points are preferentially performed in triplicate and should ideally vary less than 10%. Experiments are repeated a 2-4 times. Data points are normalized for differences in transfection efficiency using β-galactosidase, and plotted as relative activity where the untreated reporter is defined to have an activity of 1 unit.

Neither the GAL4 DBD, nor the GAL4-L66 chimera should be capable of stimulating transcription from a reporter construct containing the GAL4 upstream activating sequence. Similarly, a fusion protein containing the Herpes virus VP16 transactivation domain linked to the RXRα-LBD (VP-RXR) should be inactive when expressed alone or with the GAL4 DBD. However, when GAL4-L66 and VP-RXR are coexpressed, the reporter could be activated, indicating that L66 and RXRα interact efficiently in cells. Using similar VP16-LBD fusion proteins, interaction can be tested between L66 and receptors for peroxisome proliferators/fatty acids (PPAR), vitamin D_3 (VDR), thyroid hormone, (T_3R), retinoic acid (RAR), or other members of the nuclear receptor superfamily (see also US Pat. 6,005,086).

FIGURE CAPTIONS:

FIG. 1:

Fig. 1 shows the cDNA sequence of the L66 gene according to the invention. It also shows the reverse complement thereof.

FIG. 2:

Fig. 1 shows the protein sequence of the L66 gene according to the invention.

FIG. 3:

Fig. 3 shows the domain composition of FXR-β (L66) comprising a so called DNA binding domain "DBD" and a ligand binding domain "LBD" which are present in members of the nuclear receptor superfamily. Also shown are the respective probability values for the presence of these domains which as can be seen in the figure are extremely high.

FIG. 4:

Fig. 4 shows the cDNA sequence from L66 from Mus musculus (SEQ ID NO. 17) as well as its reverse complement (SEQ ID NO. 18) and the protein sequence of L66 from Mus musculus (SEQ ID NO. 19).

FIG. 5:

Fig. 5 shows the DNA sequence (NC Fragment) of a splice variant of L66 (SEQ ID NO. 4). It also shows the reverse complement of the sequence (SEQ ID NO. 5).

FIG. 6:

Fig. 6 shows the DNA sequence (PolyA1 Fragment, C-terminal end) of a splice variant of L66. It also shows the reverse complement of the sequence. This differs from the NC Fragment in so far as the poly-A-tail follows the exons 7 and 7B directly (SEQ ID NO. 7).

FIG. 7:

Fig. 7 shows the DNA sequence (PolyA2 Fragment, C-terminal end) of a splice variant of L66. It also shows the reverse complement of the sequence. This differs

from the NC Fragment in so far as the poly-A-tail follows the exons 7 and 7B directly (SEQ ID NO. 9).

FIG. 8:

Fig. 8 shows the DNA sequence (GF Fragment, C-terminal end) of a splice variant of L66. It also shows the reverse complement of the sequence. This differs from the NC Fragment in so far as exon 8 is not present (SEQ ID NO. 11).

FIG. 9:

Fig. 9 shows the DNA sequence (SF Fragment, C-terminal end) of a splice variant of L66. It also shows the reverse complement of the sequence. This differs from the NC Fragment in so far as exon 7 and 8 are not present (SEQ ID NO. 13).

FIG. 10:

Fig. 10 is a schematic representation of the exon distribution within the splice variants found.

FIG. 11:

Fig. 11 shows the result of a expression test on various cell lines and tissues.

FIG. 12:

Fig. 12 shows the genomic DNA sequence for L66 from Mus musculus (SEQ ID NO. 20).

FIG. 13:

Fig. 13 shows the reverse complement of the genomic DNA sequence for L66 from Mus musculus (SEQ ID NO. 21).

FIG. 14:

Fig. 14 shows a splice variant cDNA sequence of L66 from Mus musculus (SEQ ID NO. 22), its reverse complement (SEQ ID NO. 23) and the corresponding protein sequence (SEQ ID NO. 24).

FIG. 15:

Fig. 15 lists all the sequences according to the invention, their respective origin as well as in which figures they are depicted.

FIG. 16:

Fig. 16 shows a BLASTP (Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements; Nucleic Acids Res. 2001 Jul 15;29(14):2994-3005.) - alignment of L66 against tremblnew (sequence identifier |AF384555|AF384555) "NR1H4"; product: "farnesol receptor" *from Homo sapiens farnesol receptor* (NR1H4) *mRNA, complete cds, alternatively* spliced. //:gp|AF384555|14326451 gene: "NR1H4"; product: "farnesol receptor"; Homo sapiens farnesol receptor (NR1H4) mRNA, complete cds, alternatively spliced. The hit is scoring at : 7e-88 (expectation value) for an alignment length (overlap) : 389 having 46 % identities (Scoring matrix : BLOSUM62 (used to infer consensus pattern).

FIG. 17:

Fig. 17 A shows the intron – exon composition of the mouse L66 gene (numericals).

Fig. 17 B shows the intron – exon composition of spliceform variant 1 of the mouse L66 gene (numericals).

FIG. 18:

Fig. 18 A shows the intron – exon composition of the mouse L66 gene (sequences).

Fig. 18 B shows the intron – exon composition of spliceform variant 1 of the mouse L66 gene (sequences).

FIG. 19:

Fig. 19 shows a multiple sequence alignment of the L66 C4 zinc finger domain and shows the conservation of this domain.

FIG. 20:

Fig. 20 19 shows a multiple sequence alignment of the L66 protein and demonstrates the conservation of the L66 Ligand Binding Domain (LBD).

CLAIMS:

1. A nucleic acid molecule coding for a nuclear receptor which is selected from the group comprising:

   a) the nucleotide sequences set forth in SEQ ID NOs: 1, 4, 6, 8, 10, 12, 17, 20 and/or 22;

   b) or complement thereof as set forth in SEQ ID NOs: 2, 5, 7, 9, 11, 13, 18, 21 and/or 23;

   c) a nucleic acid which hybridizes to a nucleic acid having a nucleotide sequence which is the complement of the nucleotide sequence of SEQ ID NOs: 1, 4, 6, 8, 10, 12, 17, 20 and/or 22 under conditions of high stringency, and

   d) a nucleic acid which hybridizes to a nucleic acid having a nucleotide sequence which is the complement of the nucleotide sequence of SEQ ID NOs: 2, 5, 7, 9, 11, 13, 18, 21 and/or 23 under conditions of high stringency.

2. The nucleic acid molecule of claim 1 which is genomic DNA.

3. The nucleic acid molecule of claim 1 which is cDNA.

4. The nucleic acid molecule of claim 1 which is RNA.

5. A nucleic acid molecule comprising the nucleic acid molecule of any of claims 1 to 4 and a label attached thereto.

6. A vector comprising the nucleic acid molecule of claim 1.

7. The vector of claim 6, which is an expression vector.

8. A host cell transfected with the vector of claim 6 or 7.

9. A host cell transfected with the expression vector of claim 7.

10. A method of producing a polypeptide comprising the step of culturing the host cell of claim 9 in an appropriate culture medium to, thereby, produce the polypeptide.

11. An isolated polypeptide encoded by any portion of the nucleic acid of claim 1.

12. A polypeptide selected from the group comprising:

the amino acid sequences set forth in SEQ ID NO: 3, 24 and/or 19.

13. A method for screening for agents which are capable of inhibiting the cellular function of the nuclear receptor L66 comprising the steps of:

a) contacting one or more candidate agents with a polypeptide according to claim 11 or 12
b) removing unbound agent(s)
c) detecting whether the agent(s) interact with the polypeptide of the nuclear receptor.

14. A method for inhibiting the cellular function of the nuclear receptor L66, comprising the steps of:

a) contacting a cell with a binding agent of a polypeptide according to claim 11 or 12,
whereby the cellular function of L66 is inhibited.

15. Method according to claim 14,
characterized in that the binding agent is an antibody.

16. Method according to claim 14,
characterized in that the binding agent is RNA.

17. Method according to claim 14,

characterized in that the binding agent is an anti-sense oligonucleotide.

18. Method according to claim 14,
characterized in that the binding agent is a ribozyme.

19. Method according to claim 14,
characterized in that the cell is in a body.

20. Use of a nucleic acid or protein sequence according to SEQ ID NO.: 1, 4, 6, 8, 10, 12, 17, 20, 22, 2, 5, 7, 9, 11, 13, 18, 21, 3, 19, 24 and/or 23 for the construction of multiple nuclear receptor specific sequence alignments.

21. Use of the sequences according to claim 20 for the construction of protein sequence alignments.

Fig. 1

SEQ ID NO. 1.: (cDNA Sequence Homo sapiens L66)

```
5'-
cctggaataa aaaggtccag accaacctat tcttcctcga gaaataaggg acaggaagaa     60
ttctgtgtag tttgtggtga taaagcatca ccatcaccat atcattataa tgcacttacc    120
tgtgaaggtt gcaaagaaat acctatggta aaaaatttta aacttttttt attgggtttt    180
tttcaatgta gcatcnnnca aaatgcagta tatagttgca ggaatggtag tcactgtgaa    240
atggacatgt acatgcgtag aaaatgtcaa gagtgcagac tgaaaaagta taaggcagta    300
ggaatgttgg cagaatgttt gctcacagaa atccaatgta aattaaagag acttcaaaag    360
aactttaagg agaagaatca tttttactct aacatcaaag tggaagagga aggagtagac    420
cacagttttc tatcatccac cactagacct ggaaaagaaa gcatggaact aactgaagag    480
gaacatcagc tcattaataa cattgtggct gctcatcaaa aatataccat tcctttagaa    540
gaaacaaatt tgtatctgca ggaacataca aatcctgaac tgagcttttt gcaactctca    600
gagacagcag tcctacacat acgtgggcta atgaatttta ccaaggggct cccaggattt    660
gaaaatttgg ccaatgagga tcaaactgca ctacagaagg gatcaaaaac tgaagtgata    720
tttctccatg ggcccaact ttacaataca atgataattt ccatatgttt gattctaccc     780
tatgtttgga tgaaaataca ttttcgtatc agtttttttgg gtgttactga agaatttatt   840
acannnctgt tttacttcta caaaagaatg agcaaacttg atgtaactaa tactgaatat    900
gctctgcttg cagcaacaat tgttttttca gatcgtccat gccttaaaaa taagcaatat    960
atggaaaatt tannngaacc agttttacaa atattgtata agtattcaaa aatgtatcat   1020
ccagaagacc cannncattt tgcccatctc atatggaagc atactgaact gagaactctg   1080
aattataacc attcagaaat acttagcact tggaaaacaa aggaccccaa attggctact   1140
ttactctctg ag   - 3'                                                1152
```

SEQ ID NO. 2: (cDNA Reverse Complement L66)

```
5'-
ctcagagagt aaagtagcca atttggggtc ctttgttttc caagtgctaa gtatttctga     60
atggttataa ttcagagttc tcagttcagt atgcttccat atgagatggg caaaatgnnn    120
tgggtcttct ggatgataca tttttgaata cttatacaat atttgtaaaa ctggttcnnn    180
taaattttcc atatattgct tatttttaag gcatggacga tctgaaaaaa caattgttgc    240
tgcaagcaga gcatattcag tattagttac atcaagtttg ctcattcttt tgtagaagta    300
aaacagnnnt gtaataaatt cttcagtaac acccaaaaaa ctgatacgaa aatgtatttt    360
catccaaaca tagggtagaa tcaaacatat ggaaattatc attgtattgt aaagttgggc    420
cccatggaga aatatcactt cagtttttga tcccttctgt agtgcagttt gatcctcatt    480
ggccaaattt tcaaatcctg ggagcccctt ggtaaaattc attagcccac gtatgtgtag    540
gactgctgtc tctgagagtt gcaaaaagct cagttcagga tttgtatgtt cctgcagata    600
caaatttgtt tcttctaaag gaatggtata tttttgatga gcagccacaa tgttattaat    660
gagctgatgt tcctcttcag ttagttccat gctttctttt ccaggtctag tggtggatga    720
tagaaaactg tggtctactc cttcctcttc cactttgatg ttagagtaaa aatgattctt    780
ctccttaaag ttcttttgaa gtctctttaa tttacattgg atttctgtga gcaaacattc    840
tgccaacatt cctactgcct tatacttttt cagtctgcac tcttgacatt ttctacgcat    900
gtacatgtcc atttcacagt gactaccatt cctgcaacta tatactgcat tttgnnngat    960
gctacattga aaaaaaccca ataaaaaagt tttaaaattt tttaccatag gtatttcttt   1020
gcaaccttca caggtaagtg cattataatg atatggtgat ggtgatgctt tatcaccaca   1080
aactacacag aattcttcct gtcccttatt tctcgaggaa gaataggttg gtctggacct   1140
ttttattcca gg - 3'                                                  1152
```

Fig. 2

SEQ ID NO. 3: (Protein L66 Homo Sapiens)

```
PGIKRSRPTY SSSRNKGQEE FCVVCGDKAS PSPYHYNALT CEGCKEIPMV KNFKTFLLGF    60
FQCSIXQNAV YSCRNGSHCE MDMYMRRKCQ ECRLKKYKAV GMLAECLLTE IQCKLKRLQK   120
NFKEKNHFYS NIKVEEEGVD HSFLSSTTRP GKESMELTEE EHQLINNIVA AHQKYTIPLE   180
ETNLYLQEHT NPELSFLQLS ETAVLHIRGL MNFTKGLPGF ENLANEDQTA LQKGSKTEVI   240
FLHGAQLYNT MIISICLILP YVWMKIHFRI SFLGVTEEFI TXLFYFYKRM SKLDVTNTEY   300
ALLAATIVFS DRPCLKNKQY MENLXEPVLQ ILYKYSKMYH PEDPXHFAHL IWKHTELRTL   360
NYNHSEILST WKTKDPKLAT LLSE                                          384
```

Fig. 3

Domain composition:

Domain	Region from protein	E-value
ZnF_C4	19-105	14.8e-15 (A below)
HOLI	201-360	9.99e-13 (B below)

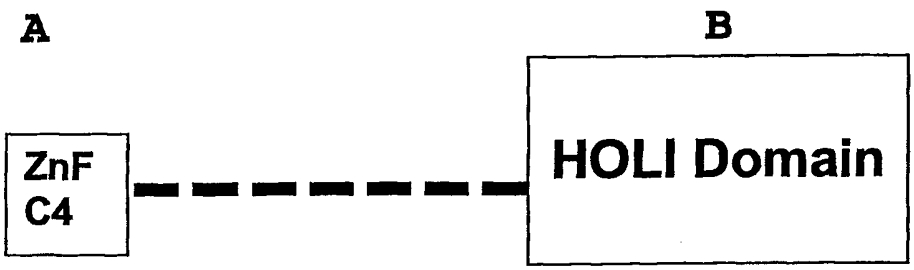

Fig. 4

cDNA L66 from Mus musculus (SEQ ID NO. 17)

```
atgttaataa aaccagatat tttgccagaa caattccatt atcagctgtg tgatacagat     60
ttccaagaac caccctattg tcaatattct accgctcagt ttcctccagc gttacagtcc    120
ccatctttac aaagtcattt caacacacat ggcttggatc cacagtacag tggaggcagt    180
tggtgtggac tcgacgctcg agaatctggt cagtccactt atgtggttgt tcacgatgat    240
gaagatgaat tccctggggc acaaaggtgc agagcaactt gttctttacg ctggaagggt    300
caagatgaca tgctctgcat ggtctgcggt gataaggcat caggatatca ctacaatgca    360
cttacttgtg aggggtgcaa aggctttttc cggcgtagca ttaccaagaa tgcagtgtat    420
tcttgcaaga acggtggtca ctgtgaaatg gacatgtaca tgcgcagaaa atgccaagag    480
tgcagactga agaagtgtaa ggcggtgggg atgttggcag aatgtttgct cacagagatc    540
cagtgtaagt caaagagact tcgcaagaac ttcaagcacg ggcctgccct gtaccctgcc    600
atccaagtgg aagatgaagg agcagacacc aaacacgtgt catccagcac cagatctggg    660
aaaggggttc aggacaacat gactctaact caagaggaac atcggcttct gaataccata    720
gtgactgctc accaaaaatc catgattccc ttgggagaaa caagcaaact tctgcaggag    780
ggttccaacc ccgaactaag ttttctgaga ctctcagagg tatcagtcct gcacatacaa    840
gggctaatga agtttaccaa gggactccca ggatttgaaa atttaaccac tgaggatcag    900
gctgcattac agaaggcgtc aaaaactgaa gtgatgttcc ttcatgtagc ccagctttat    960
ggtgggaaag actcaacctc tggaagtact atgagaccag caaagccctc agctgggaca   1020
ctagaggtgc ataatcctag cgctgatgaa agtgttcatt ctccggaaaa ctttctcaag   1080
gaaggctacc cttcggctcc tctaactgat attactaaag aatttattgc ctcactatct   1140
tacttctaca gaagaatgag tgaacttcat gtatcggata ctgaatatgc tctgcttacg   1200
gcgacaacag tgcttttctc agatcgtcca tgccttaaaa ataagcagca tatagaaaac   1260
ctacaagaac cagtcctgca acttttgttt aagttttcaa aaatgtacca tccagaagac   1320
ccacagcatt tcgcccacct catagggagg cttactgaac tgagaactct gagtcacagc   1380
cactctgaaa tccttcgcat gtggaaaaca aaggacccca ggttggtgat gttattctct   1440
gagaaatggg atctgcactc attttcctga                                     1470
```

Reverse Complement of cDNA L66 from Mus musculus (SEQ ID NO. 18)

```
tcaggaaaat gagtgcagat cccatttctc agagaataac atcaccaacc tggggtcctt     60
tgttttccac atgcgaagga tttcagagtg gctgtgactc agagttctca gttcagtaag    120
cctccctatg aggtgggcga aatgctgtgg gtcttctgga tggtacattt ttgaaaactt    180
aaacaaaagt tgcaggactg gttcttgtag gttttctata tgctgcttat ttttaaggca    240
tggacgatct gagaaaagca ctgttgtcgc cgtaagcaga gcatattcag tatccgatac    300
atgaagttca ctcattcttc tgtagaagta agatagtgag gcaataaatt ctttagtaat    360
atcagttaga ggagccgaag ggtagccttc cttgagaaag ttttccggag aatgaacact    420
ttcatcagcg ctaggattat gcacctctag tgtcccagct gagggctttg ctggtctcat    480
agtacttcca gaggttgagt ctttcccacc ataaagctgg gctacatgaa ggaacatcac    540
ttcagttttt gacgccttct gtaatgcagc ctgatcctca gtggttaaat tttcaaatcc    600
tgggagtccc ttggtaaact tcattagccc ttgtatgtgc aggactgata cctctgagag    660
tctcagaaaa cttagttcgg ggttggaacc ctcctgcaga agtttgcttg tttctcccaa    720
gggaatcatg gatttttggt gagcagtcac tatggtattc agaagccgat gttcctcttg    780
agttagagtc atgttgtcct gaaccccttt cccagatctg gtgctggatg acacgtgttt    840
ggtgtctgct ccttcatctt ccacttggat ggcagggtac agggcaggcc cgtgcttgaa    900
gttcttgcga agtctctttg acttacactg gatctctgtg agcaaacatt ctgccaacat    960
ccccaccgcc ttacacttct tcagtctgca ctcttggcat tttctgcgca tgtacatgtc   1020
catttcacag tgaccaccgt tcttgcaaga atacactgca ttcttggtaa tgctacgccg   1080
gaaaaagcct ttgcacccct cacaagtaag tgcattgtag tgatatcctg atgccttatc   1140
accgcagacc atgcagagca tgtcatcttg acccttccag cgtaaagaac aagttgctct   1200
gcacctttgt gccccaggga attcatcttc atcatcgtga acaaccacat aagtggactg   1260
accagattct cgagcgtcga gtccacacca actgcctcca ctgtactgtg gatccaagcc   1320
atgtgtgttg aaatgacttt gtaaagatgg ggactgtaac gctggaggaa actgagcggt   1380
agaatattga caatagggtg gttcttggaa atctgtatca cacagctgat aatggaattg   1440
ttctggcaaa atatctggtt ttattaacat                                     1470
```

Fig. 4 continu d

Protein L66 from Mus musculus (SEQ ID NO. 19)

```
MLIKPDILPE QFHYQLCDTD FQEPPYCQYS TAQFPPALQS PSLQSHFNTH GLDPQYSGGS    60
WCGLDARESG QSTYVVVHDD EDEFPGAQRC RATCSLRWKG QDDMLCMVCG DKASGYHYNA   120
LTCEGCKGFF RRSITKNAVY SCKNGGHCEM DMYMRRKCQE CRLKKCKAVG MLAECLLTEI   180
QCKSKRLRKN FKHGPALYPA IQVEDEGADT KHVSSSTRSG KGVQDNMTLT QEEHRLLNTI   240
VTAHQKSMIP LGETSKLLQE GSNPELSFLR LSEVSVLHIQ GLMKFTKGLP GFENLTTEDQ   300
AALQKASKTE VMFLHVAQLY GGKDSTSGST MRPAKPSAGT LEVHNPSADE SVHSPENFLK   360
EGYPSAPLTD ITKEFIASLS YFYRRMSELH VSDTEYALLT ATTVLFSDRP CLKNKQHIEN   420
LQEPVLQLLF KFSKMYHPED PQHFAHLIGR LTELRTLSHS HSEILRMWKT KDPRLVMLFS   480
EKWDLHSFS                                                          489
```

Fig. 5

L66 NC Fragment

```
cctggaataa aaaggtccag accaacctat tcttcctcga gaaataaggg acaggaagaa   60
ttctgtgtag tttgtggtga taaagcatca ccatcaccat atcattataa tgcacttacc  120
tgtgaaggtt gcaaagcatc aacaaaatgc agtatatagt tgcaggaatg gtagtcactg  180
tgaaatggac atgtacatgc gtagaaaatg tcaagagtgc agactgaaaa agtataaggc  240
agtaggaatg ttggcagaat gtttgcacac agaaatccaa tgtaaattaa agagacttca  300
aaagaacttt aaggagaaga atcattttta ctctaacatc aaagtggaag aggaaggagt  360
agaccacagt tttctatcat ccaccactag acctggaaaa gtgattcagg aaagcatgga  420
actaactgaa gaggaacatc agctcattaa taacattgtg gctgctcatc aaaaatatac  480
cattccttta gaagaaacaa atttctgcag gaacatacaa atcctgaact gagctttttg  540
caactctcag agacagcagt cctacacata cgtgggctaa tgaattttac caaggggctc  600
ccaggatttg aaaatttggc caatgaggat caaactgcac tacagaaggg atcaaaaact  660
gaagtgatat ttctccatgg ggcccaactt tacagtcaga aacaatcagc ctctgaaagt  720
tctgtgagaa tattaaatca ttcagattat acaccaaatt gtcacaatag gagtggtgat  780
agaagtctta tttgttctat ggaaaaattt tacaatgaag aatgtccttc tactactcta  840
attggtttca agaagctcat ctggaaaatg gtgataataa tagaacttac ctcatacagt  900
attgtgacta ctacataaaa taatacatag agatcgtcca tgccttaaaa ataagcaata  960
tatggaaaat ttacaagaac cagttttaca aatattgtat aagtattcaa aaatgtatca 1020
tccagaagac ccatagcatt ttgcccatct catatggaag catactgaac tgagaac    1077
```

L66 NC Fragment (Reverse Complement)

```
gttctcagtt cagtatgctt ccatatgaga tgggcaaaat gctatgggtc ttctggatga   60
tacatttttg aatacttata caatatttgt aaaactggtt cttgtaaatt ttccatatat  120
tgcttatttt taaggcatgg acgatctcta tgtattattt tatgtagtag tcacaatact  180
gtatgaggta agttctatta ttatcaccat tttccagatg agcttcttga aaccaattag  240
agtagtagaa ggacattctt cattgtaaaa tttttccata gaacaaataa gacttctatc  300
accactccta ttgtgacaat ttggtgtata atctgaatga tttaatattc tcacagaact  360
ttcagaggct gattgtttct gactgtaaag ttgggcccca tggagaaata tcacttcagt  420
ttttgatccc ttctgtagtg cagtttgatc ctcattggcc aaattttcaa atcctgggag  480
ccccttggta aaattcatta gcccacgtat gtgtaggact gctgtctctg agagttgcaa  540
aaagctcagt tcaggatttg tatgttcctg cagaaatttg tttcttctaa aggaatggta  600
tatttttgat gagcagccac aatgttatta atgagctgat gttcctcttc agttagttcc  660
atgctttcct gaatcacttt tccaggtcta gtggtggatg atagaaaact gtggtctact  720
ccttcctctt ccactttgat gttagagtaa aaatgattct tctccttaaa gttcttttga  780
agtctcttta atttacattg gatttctgtg tgcaaacatt ctgccaacat tcctactgcc  840
ttatactttt tcagtctgca ctcttgacat tttctacgca tgtacatgtc catttcacag  900
tgactaccat tcctgcaact atatactgca ttttgttgat gctttgcaac cttcacaggt  960
aagtgcatta taatgatatg gtgatggtga tgctttatca ccacaaacta cacagaattc 1020
ttcctgtccc ttatttctcg aggaagaata ggttggtctg gaccttttta ttccagg    1077
```

Fig. 6

L 66 PolyA1 Fragment

```
cccaacttta cagtcagaaa caatcagcct ctgaaagttc tgtgagaata ttaaatcatt    60
cagattatac accaaattgt cacaatagga gtggtgatag aagtcttatt tgttctatgg   120
aaaaatttta caatgaagaa tgtccttcta ctactctaat tggtaatatg actcaatatg   180
aaatattata ttggatgcaa aaatttgtat ataatgttta actttcttat actgctttga   240
gatacaatga taatttccat atgtttgatt ctaccctatg tttggatgaa aatacatttt   300
cgtatcngcc caaaaaaaaa aaaaaaaaaa a                                  331
```

L 66 PolyA1 Fragment (Reverse Complement)

```
tttttttttt tttttttttt gggcngatac gaaaatgtat tttcatccaa acatagggta    60
gaatcaaaca tatggaaatt atcattgtat ctcaaagcag tataagaaag ttaaacatta   120
tatacaaatt tttgcatcca atataatatt tcatattgag tcatattacc aattagagta   180
gtagaaggac attcttcatt gtaaaatttt tccatagaac aaataagact tctatcacca   240
ctcctattgt gacaatttgg tgtataatct gaatgattta atattctcac agaactttca   300
gaggctgatt gtttctgact gtaaagttgg g                                  331
```

Fig. 7

L 66 PolyA2 Fragment

```
cccaacttta cagtcagaaa caatcagcct ctgaaagttc tgtgagaata ttaaatcatt    60
cagattatac accaaattgt cacaatagga gtggtgatag aagtcttatt tgttctatgg   120
aaaaatttta caatgaagaa tgtccttcta ctactctaat tggtaatatg actcaatatg   180
aaatattata ttggatgcaa aaatttgtat ataatgttta actttcttat actgctttga   240
gatacaatga taatttccat atgtttgatt ctaccctatg tttggatgag gaaaaaaaaa   300
aaaaaaaaaa a                                                        311
```

L 66 PolyA2 Fragment (Reverse Complement)

```
tttttttttt tttttttttt cctcatccaa acatagggta gaatcaaaca tatggaaatt    60
atcattgtat ctcaaagcag tataagaaag ttaaacatta tatacaaatt tttgcatcca   120
atataatatt tcatattgag tcatattacc aattagagta gtagaaggac attcttcatt   180
gtaaaatttt tccatagaac aaataagact tctatcacca ctcctattgt gacaatttgg   240
tgtataatct gaatgattta atattctcac agaactttca gaggctgatt gtttctgact   300
gtaaagttgg g                                                        311
```

Fig. 8

L 66 GF Fragment (DNA Sequence)

```
ggccaatgag gatcaaactg cactacagaa gggatcaaaa actgaagtga tatttctcca    60
tggggcccaa ctttacagtc agaaacaatc agcctctgaa agttctgtga gaatattaaa   120
tcattcagat tatacaccaa attgtcacaa taggagtggt gatagaagtc ttatttgttc   180
tatggaaaaa ttttacaatg aagaatgtcc ttctactact ctaattgatc gtccatgcct   240
taaaaataag caatatatgg aaaatttaca agaaccagtt ttacaaatat tgtataagta   300
ttcaaaaatg tatcatccag aagacccata gcattttgcc catctcatat ggaagcatac   360
tgaactgaga actctgaatt ataaccattc agaaatactt agcacttgga aaacaaagga   420
cccca                                                               425
```

L 66 GF Fragment (DNA Sequence Reverse Complement)

```
tggggtcctt tgttttccaa gtgctaagta tttctgaatg gttataattc agagttctca    60
gttcagtatg cttccatatg agatgggcaa aatgctatgg gtcttctgga tgatacattt   120
ttgaatactt atacaatatt tgtaaaactg gttcttgtaa attttccata tattgcttat   180
```

```
ttttaaggca tggacgatca attagagtag tagaaggaca ttcttcattg taaaatttt
ccatagaaca aataagactt ctatcaccac tcctattgtg acaatttggt gtataatctg     300
aatgatttaa tattctcaca gaactttcag aggctgattg tttctgactg taaagttggg     360
ccccatggag aaatatcact tcagtttttg atcccttctg tagtgcagtt tgatcctcat     420
tggcc                                                                  425
```

Fig. 9

L 66 SF Fragment (DNA Sequence)

```
ggccaatgag gatcaaactg cactacagaa gggatcaaaa actgaagtga tatttctcca    60
tggggcccaa ctttacagtc agaaacaatc agcctctgaa aatcgtccat gccttaaaaa   120
taagcaatat atggaaaatt tacaagaacc agttttacaa atattgtata agtattcaaa   180
aatgtatcat ccagaagacc catagcattt tgcccatctc atatggaagc atactgaact   240
gagaactctg aattataacc attcagaaat acttagcact tggaaaacaa aggacccca    299
```

L 66 SF Fragment (DNA Sequence - Reverse Complement)

```
tggggtcctt tgttttccaa gtgctaagta tttctgaatg gttataattc agagttctca    60
gttcagtatg cttccatatg agatgggcaa aatgctatgg gtcttctgga tgatacattt   120
ttgaatactt atacaatatt tgtaaaactg gttcttgtaa attttccata tattgcttat   180
ttttaaggca tggacgattt tcagaggctg attgtttctg actgtaaagt tgggccccat   240
ggagaaatat cacttcagtt tttgatccct tctgtagtgc agtttgatcc tcattggcc    299
```

Fig. 10:

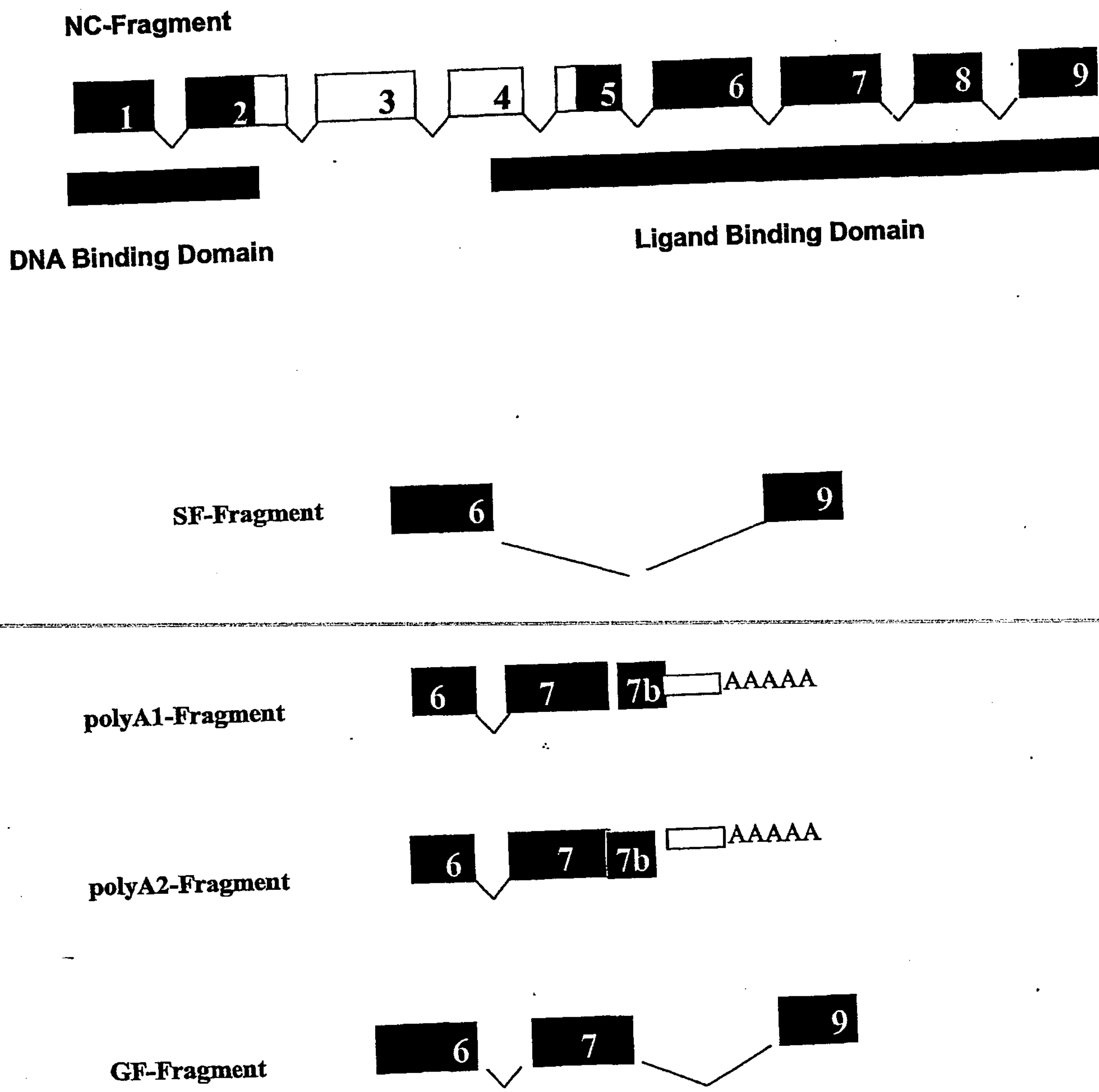
NC-Fragment
1
2
3
4
5
6
7
8
9
DNA Binding Domain
Ligand Binding Domain
SF-Fragment
6
9
polyA1-Fragment
6
7
7b
AAAAA
polyA2-Fragment
6
7
7b
AAAAA
GF-Fragment
6
7
9

Fig. 11:

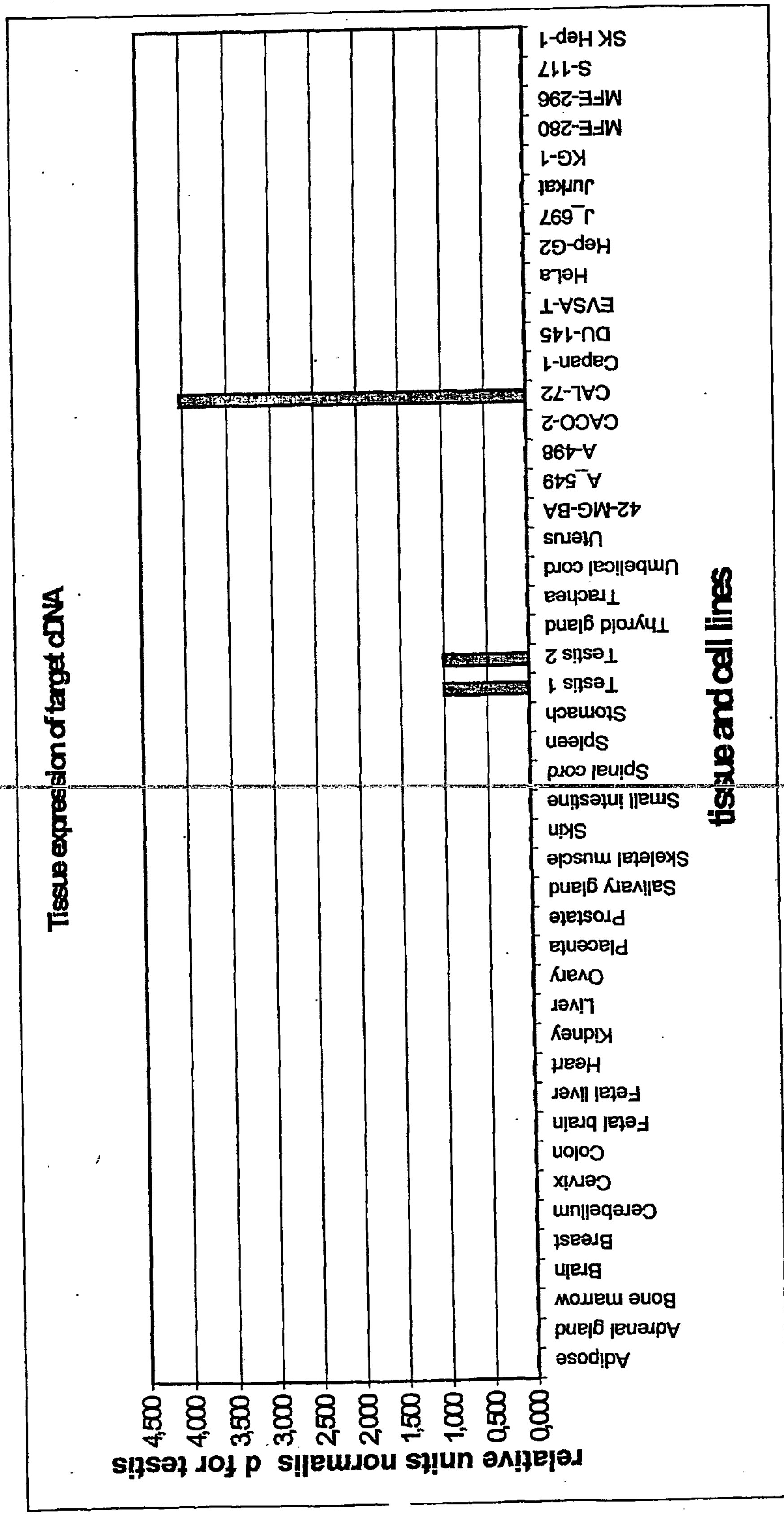
Tissue expression of target cDNA
relative units normalis d for testis
4,500
4,000
3,500
3,000
2,500
2,000
1,500
1,000
0,500
0,000
tissue and cell lines
Adipose
Adrenal gland
Bone marrow
Brain
Breast
Cerebellum
Cervix
Colon
Fetal brain
Fetal liver
Heart
Kidney
Liver
Ovary
Placenta
Prostate
Salivary gland
Skeletal muscle
Skin
Small intestine
Spinal cord
Spleen
Stomach
Testis 1
Testis 2
Thyroid gland
Trachea
Umbelical cord
Uterus
42-MG-BA
A_549
A-498
CACO-2
CAL-72
Capan-1
DU-145
EVSA-T
HeLa
Hep-G2
J_697
Jurkat
KG-1
MFE-280
MFE-296
S-117
SK Hep-1

Fig. 12:

```
ctttggggtc atagcattta ttaggtattg ttagttccat ttactgtgtt cctattgtgc    60
tggacggcaa gtcctgagat ggaatagatg tgttgaacta actttctgtt gagtataaaa   120
cattttgagg tttaaaacaa ttatgaactt ttttgacctt ttatgatttg ttcttttttta   180
tgttaataaa accagatatt ttgccagaac aattccatta tcagctgtgt gatacagatt   240
tccaagaacc accctattgt caatattcta ccgctcagtt tcctccagcg ttacagtccc   300
catctttaca aagtcatttc aacacacatg gcttggatcc acagtacagt ggaggcagtt   360
ggtgtggact cgacgctcga gaatctggtc agtccactta tgtggttgtt cacgatgatg   420
aagatgaatt ccctggggca caaaggtgca gagcaacttg ttctttacgc tggaagggtc   480
aagatgacat gctctgcatg gtctgcggtg ataaggcatc aggatatcac tacaatgcac   540
ttacttgtga ggggtgcaaa ggtaagggta aggcgcaacc aaacagacag tgctgcgcac   600
gcgcacacgt gtgtatgtgt gtgtgcacgt gcgtgtgcaa tcgttataaa acatgagaaa   660
caaaaccagg aaaaagtatg tctaaaggga tcaatctgtg tgatcaaaga taggagaata   720
aaatgttaag tttaaggtta gtttctttcc tttcctttcc tttcctttcc tttcctttcc   780
tttcctttcc ctttcccttt ccctttccct ttcccttcct ctttctttct ttctttcttt   840
ctttctttct ttctttcttt ctttctttct ttctttcttt ctttctttct ctctctctct   900
ctctctctct ctctctctct ctctctctct ctctctctct ctccctctct ctctctctct   960
ctctctctct ctctctctct ctctctctct ctctctctct ctctctctct ctctctccct  1020
ccctcccccc ctctctctct tttgaagtag agtatcattg tatagccctg gctggcctag  1080
acctcactat ggagatcagg ttggctttga actcacagat acctgcctgt ctctgctttc  1140
taggtgctgg gattaaaggt gtgggccacc atgcctggac catttctttc tttatttttt  1200
tttctcctcc cttgtcttcc ccttcgtcct catcttctcc tcctcatctt tcttcttttt  1260
ttgagaaaag gtctctctat gcatccctgg ctgtcctgga actcactgtt tagaccaggc  1320
tgtccttgaa ctctcagaga tccacctacc tttgtatcct tactgctggc attaagagca  1380
ttcactacca ggtctgtctc ttcatttttg atgaagactt tcacacttta ctttaaactc  1440
gatacatata agcaaacgtt ttccctattt ctagaagaat ttgatgctgg aacaaaacat  1500
gacaatgtaa agatctgtaa accatgacac ctttttatgt ggttgataaa atttatacac  1560
ttggagaaaa tttgaaataa atgacttcta tgattctaag tcaataaatt ttaaaaagtg  1620
tttcgtatag aaagagaaca agctgaaagt cgggaagcaa gggccagtgt ttgcaatgat  1680
atttgtaaag agcggaagcc tctctgataa ataacagttc tgaaacagtc taccactcca  1740
gaaatggata aacatcggga tgggaatgca ggccagtggt agagcactta gcacctactt  1800
ttggatttga ttctagcacc ctctgaaaca aaacaaaaca acccaaacaa cagaaaacag  1860
tttagtcaat gtaacacaaa taaacgtgtt tgtttagggt aaaatttac ctgtaaatct  1920
taaacctcaa cactggcaaa gaatagccaa aggatatgag gtgacagagt cctgtggcga  1980
gtagttaggc aggggtggcc tgctcttgac ggtgacgggt tactgttaaa actgagctga  2040
agaacccagc acagtttga aaacctgagc tgagcttttt cttccttcgt tgaactgaaa  2100
tggcgaaact ccatctttca agtgtccact gcaagagagt aagtggtggg gctttgggac  2160
aactcatagg agggagagca ggaaatggca ggtattactg tgtaataagg ttcgaaggag  2220
gaaaagggat gggaactatc tggcaagtgc tgggaaaggg agagagagca tctgcggatg  2280
agggagggag ggaccacctt tgtgggctgc aactcagact gggagggttt agaaaagaga  2340
ggcttttctg aattttttct ttattttagg tgaccatgta ttctctcaaa ggaaactcga  2400
gagatgcatt gtaggatctt gatattttgt ataatgcagt taaaaaaaaa attcttatag  2460
gtaatttacc tatagtaggg tagagagttg tcttaatgtg ataatgatac tccatttaat  2520
ctccagacgt tataaagaaa gacccagcct caggatgtca actgaactat actaactata  2580
aagaattata gaattagcta actttccaga taaatgtata tttaaaaaaa tgtcgggaat  2640
aaaaaaaacc ctatttaatt gaaatactct agttagcttg attactttct ataaaggtaa  2700
aatgttaata aaaatgagaa gtaactaaac atatccatgt caaactttga ccatctttgt  2760
tgttaggctt tttccggcgt agcattacca agaatgcagt gtattcttgc aagaacggtg  2820
gtcactgtga aatggacatg tacatgcgca gaaaatgcca agagtgcaga ctgaagaagt  2880
gtaaggcggt ggggatgttg gcagaatgta agtgccagtg tgttcgtggc ttgcttttct  2940
gaggggattt aacttttccc agccttacct tcctgacaag gaggacatga tagtggtggt  3000
ggtttttcca gggcaagaat cgcccattgc actagggatg tgctgactga cccctgtgat  3060
ttccacaaat gtgggaaact caggtgcatc atttgtgtta gtgaaggact gtattcacac  3120
tctcctctat ataataaaaa tgaatttaaa aattaataaa aaattaaaaa tttccaacct  3180
tcttgaggac ataaagcatt cttcaagagt tattttgaac atggccgtat tcgtttgtat  3240
atgcacacca cacacatatg tacagtaatt gcgccacgtg gatggcccat tatcaattta  3300
agtagtgatt cccataaatg ttgccctgtt ctatattgct ctaattttga aatcagtcat  3360
agtttcatct taattagatg ttctctgaga cctgaagagt agttaggcag ttaagagcat  3420
taggagagga cctgggcccg gtccccagct tccacgtggc ccctcacaat catttataac  3480
tcagttcctg agcatcccag gccctcttct gacctctgca ggtatgagtg ttgtgtgcat  3540
```

Fig. 12 continu d:

acacacatgc aagcaaaaca cacacataaa ataaaagaaa cctccatagt gatttaaaaa 3600
aggtattgta tgtttctgtt acttaaccta aaaatatata ctgccttgaa aagtgcaaag 3660
ttaaagttag tagttcccaa aaagaaagaa agaggaggag aaagaagaga aaaggaagtc 3720
ttccctttct gagaaggggt tctcagagtc ctaacacact acagtcttag gacaattgaa 3780
gggtaactgg atctctggac ccctgacatc ctcatatgta aagtccacta atagtctcac 3840
gagtgtcagg actcatctct gaatatcaac acattagcat atgagcagtg ggagatacaa 3900
ctcagtccaa agaaaccttt gaaaaagttg tcttttattt tattcattaa tttaatttgt 3960
attcaggatc tcactatgta gctgacctgg aacttgctat gtacgccagg ctggcatcaa 4020
actcagagat ccccctgcct ctgccctcca atttctggct caaaagttgt cttttaacct 4080
gtggcaggca atgcatgtga aagggaagtt ggagtcctct gaggaactct atgcataaaa 4140
tgagtcatcc tacacttaaa gaaacttaaa gaaattctgg caaggtgaag ctcgagagag 4200
ttggtgagca gtcacggtga ttggcgagca taagctggtc cgcaccctgc caccattgct 4260
tggatctgca cacttcattt acctttcagt tgtcctaata ctgccatgtg ttagtgctct 4320
gacggcccct cctcctcttc cttccctcat tctcctttcc ctttcctcct ttcttttcct 4380
cttcctcctc ttcccctct tcttcctcct tttcctcctc ttcgttctct tcttgtggtt 4440
ttccttctca tggtcctggg ggttgagttc aggtttctca catcctagac aatctaccac 4500
taagctacac ccatagactc ctttcctact aaaaaacact gagttgccat ggctggtatt 4560
caacgtgaga tcatcttgta acagtgtctt aagtagctgg ggttataggt gtttactacc 4620
atgttcaata ctcaatatgc ttaaaatggc tttatttttt tgtgtattct gttctctgac 4680
aaataataaa ataatataga ggggttttaa ggttttgtt tatacttcag ctcctcatta 4740
gtttttatt ttataggtgt gtttttaaaa atactttgaa aacgacaggt atattaatag 4800
ttggatactt ttagtagctg ctaaaagtac ttagagtagt ccagtccagt ccagtacata 4860
cagtacagct cacaaccagc tatagtctca cttcaagggg ctatgacccc tctgtgctct 4920
gtgggcacct gcactcacct gtgcatttcc ccatacagac tcatgcaagt aattataaaa 4980
tatttaaaaa caaacaaaaa caatcctaaa ataaatatta tttgcctttg aaaaagaatt 5040
aaatgctttg tctgacatag gaaaaattga cttagcttat ttttaaaaat cctaagatag 5100
tctttttaa ccttattaat aaaaccagtg ataatagtca tagataatgt aattctacta 5160
ttctatcaaa tgacttttc acttccaaga tgcaactaat gtgtaaatct cttatgtaaa 5220
aatctaaaca taatttggaa atgaaaaatt aattcttaat ctttgttata actattttta 5280
agcagacaag gtgtctatga agggaaaagc agaataggaa atttaataga tgcagacatg 5340
cacgagagat atctaggact atttactcac tgaagaccat tttaatcaaa ccctacagtg 5400
ttgttttccg agccaaaagt acttaatgtg aattctgtaa tgtcacaggt ttgctcacag 5460
agatccagtg taagtcaaag agacttcgca agaacttcaa gcacgggcct gccctgtacc 5520
ctgccatcca agtggaagat gaaggagcag acaccaaaca cgtgtcatcc agcaccagat 5580
ctgggaaagg ggcaagacgc ttcttccata cagtctcttg tgtctttatg aacagtgtgt 5640
aacaagcaca cattacaggc gcctctgtt tccttatagg ttcaggacaa catgactcta 5700
actcaagagg aacatcggct tctgaatacc atagtgactg ctcaccaaaa atccatgatt 5760
cccttgggag aaacaagcaa acttgtatgt atccattgct ggcaagaaat aatttgtaat 5820
gaatgatatt aggaaatgat agcaagaaca gtaaaaaaga aacaccttag taattatcag 5880
aaagacatac tatgaatcat ttttatttat ttattttttt tatggaccag ctgcaggagg 5940
gttccaaccc cgaactaagt tttctgagac tctcagaggt atcagtcctg cacatacaag 6000
ggctaatgaa gtttaccaag ggactcccag gtaaaattct gtttcattat tttttaagga 6060
tttatttatt tttatgttta tgtgttatgc atgttttgat tgcaattata tattcatatc 6120
atgtacacaa catattgtgg aggccagaag agagtgtttg gtcccctaga actagagttt 6180
cctaccgatg gctgtgaatt accatgtggg cgttggaacc agagtctctg caagagcaag 6240
acaggctctt aactgctgat ctgtctttct agccccccaat ttttaaaaaa gattatttat 6300
ttatttattt atttatttat ttatttattt atttatctga gtacattgta gctgtacaga 6360
tggttgtgag ccttcatgtg gttgttgaga attgaatttt taggacctct gctcactctg 6420
gtcggccccg ctcgttctgg cccaaagatt tcataagtac actgtagctg tcttcaggtg 6480
caccagaaga gggcgttaga tctcgttacg gatggttgcg agccaccgtg tggttgctgg 6540
gacttgaact cagggctttt taggaagagt agtcagtgct cttacccact gagctatctc 6600
tccagcccca aattttcttt tttttaaagg gatagtatat caaacagaaa tttctagttg 6660
tggtactaga tctatcgctc agaggcagag cacctgtcta atgtgtgtga agccctgggt 6720
tttctagcaa tgttgcttct aactaaaaaa aaaaaaaatg tttatttgtc tttgtcaatt 6780
aatcagatga gtgacaatct aaatatttac aaaaattatt ataatgatgt gggctttctt 6840
agatgaaaca agaaacaaag aaggtttcct ctataatact aaacaagtca cagtttgacc 6900
aaaacttata tataactcag ctttattata tgcttagagg tacagaaact ctttacaaaa 6960
ttacattctt caaggctaat gcagctatgt tgatttgtcc ttgctgtaag aaagtgtaac 7020
ctagtgctga taaaagtgcc ccaaaataga aaaaaagaat tttattcaga tttcaattta 7080
tatactattg ttattcccca gaagcactta gggttaataa gtaaatattt ttaaacattt 7140

Fig. 12 continu d:

aatttatgaa cgttctcttt tgattatctg tttctttatg ctttatttat ttatttattt 7200
atttattttt aaggatttga aaatttaacc actgaggatc aggctgcatt acagaaggcg 7260
tcaaaaactg aagtgatgtt ccttcatgta gcccagcttt atggtgggaa agactcaacc 7320
tctggaagta aaaagaaaaa cagaatttta ttcaacttca agtacacaca gacagacaga 7380
cagacagaca gacacacaca cacacacaca cacacacaca cacacacacg cacgcacgca 7440
cgcacacact tttaaacact gtttttgga aatagactct catgtagtcc aggctatgta 7500
gaactcagta tgtagctaag gaggatctta aagccctgct gtgtctgctc ttacctgcca 7560
accatcagcc tagagcctct atgcctgcct tctatggtgc tggggacaga agccagggtc 7620
tgtgctcagt aagccagaac tcctaactcc taacatcatc agcctttcat ttatcttttc 7680
attttatttt accagttagt gtaggaacaa atgtctcata gtagctgttt ataaatagtt 7740
tcacattcat ctagcatttt ctctctaccg tgtaattctt ttaaaaattg tatttcttat 7800
tcttgtgaga actgtataca tgtcctaagt gttttttttt ttaaatctta ctcacctcac 7860
ttccttccca ccagctcctt tagaatcccc cacacacctc tctcctaact tcatgtcctt 7920
ccctgctcct tctttttcaa aataacccat tgactctaaa tagtgctgct catatacacg 7980
ggggtggggg tagggtcatc cactgaagca tggccaactt agagggctca gttccctaaa 8040
gagactggat tgtccttccc ctagcagctg ccaacagctt ctcagctgtg ggtggggctt 8100
ctgtgggtgg ggcttctgtg ctccttagac aggctcttcc aatgaggagg acttttatgg 8160
gtgatcaaac ataaaggcgt tggtgattct aagacacaaa gcagctctgg ggacggccgc 8220
atgcttgctt tgaaactcat aaccacgggc tactgtttgt ggaacgtctc tttactgggg 8280
tgaagcccta ttaggttagt gttgaaccaa ctatgcagga cctcacttca ataaaaaaaa 8340
agcaagtatt ttttaaaaaa tcaataaacc atttgaatag agcaggaaag ccatcattgt 8400
ttagtgtgtg aatgcaagag tccaatctgt aaagtcgcat gttggttgag gacatgggaa 8460
gaggcaatag gacaaaatca aaaccaaaag ccaacaatgc tagctgcctt cactagctgt 8520
actcgttaca gtgcacagcc ttttaaaata tccttattta ttgattgttc tctgcaaaca 8580
ctagaccaga aggctctaaa gggaggccct tgatgacctg gtccatgatt gcttgagtag 8640
tgtttgctta aaagtctcca cagcctctgg tgcactgagt ggttcgtgga gggtgtgtgt 8700
gtttaagaaa gccattctta gtttctattt cattatgcaa aactcatttt ggagggtctt 8760
ctgtacaatc tttttttttt ttttacaaaa tatgaactaa tatctgtttt gtgttccagt 8820
aacaaaaatg agaatgatat cattttagga gttttttttt cccccacata atagttatgc 8880
tggatactga aaagatactg aatttactat ttatataatg attaagaaag actatcaaca 8940
tttgtgtttt tctcttttgc taggtactat gagaccagca aagccctcag ctgggacact 9000
agaggtgcat aatcctagcg ctgatgaaag tgttcattct ccggaaaact ttctcaagga 9060
aggctaccct tcggctcctc taactggtaa cactgtgcag taggaagcat caccttgggt 9120
ggacacgctt cctttctcc cctttcatt gcactacaaa gagaaggcca tgggctttct 9180
cacagctcac aggcgtctct gtggtggtga aagcatttat agacacagcc aatacagttg 9240
ttagtgttgg caggtcatca gtgagaattc tatttgtaac catttctta ctttggcttc 9300
atcatggcaa ttctttccct atttaattgc gaagccattt ggggcctctg aattattctt 9360
aggatgatga tgtaatctac agccaacatc ccagtatcaa ccacttttat gggtagaagg 9420
gtaagcctta ctgtagataa tagagataca cggagagtgg cttctgtagg ctaaaggctc 9480
tcagttccac agcctcctcc ttggtccacg tgatgacaac ctctctttgt tctggtcctt 9540
ttcagtagca cacctaatgc gcccactctt gaggttcagg cctgggcttc aagggcacaa 9600
caaaggggat tgtctgtgcc ggccatcaca tcggtagcca tcttgaaccc cactctcttt 9660
tcaattttct tacactgttt gagagaagaa aagacaatca caatgtgttt gcttctgtgc 9720
tccattccga tgaaaattca cttatattaa tcttttcaga tattactaaa gaatttattg 9780
cctcactatc ttacttctac agaagaatga gtgaacttca tgtatcggat actgaatatg 9840
ctctgcttac ggcgacaaca gtgcttttct caggtacaga ctgagaagta aggctcggag 9900
tgatctctgg acaggcaatc actaagttct ttttatcgaa ctatgaactg aagttgaatg 9960
taccactctc atcccatcat gctttgctct ctatgtgcat tcctttaagc ttaccactgg 10020
gttacatctt tcttgcatct taaagaaatg ctttcttgcc ctcatagccc ctcagatata 10080
cgctcccagt ccctgataga gagaggttct cagggagtta gctatattct agtttccacc 10140
ttactcatgc tcactcctga gttcaccaca gtcacatttc actccaccct ctccagtcca 10200
ctacaacctg tcaggatctt caggacatgc tatcactgac tcaactggtc aatgctctgt 10260
cttcatctta cttgccccac tggtgtattt tgtagagagt atcataagca tctgaaagaa 10320
aagaagggga tatgttgttt cctgtgagcc cttcagtgga attttgaagc acggtgaaga 10380
tgtctgatgt gttaggagaa tttctcatgc ttctttatgt cagttcacag tcttcttgac 10440
tatgcgttac aaaccggacc caggttgaag tgagacagtt acttagcatg ggtaacccat 10500
tgctcaagcc agcgacgtct tcctagggga ccattgtcct gtgtccttca ccgattcggc 10560
catgtttgtg ttcttttata ttctaacttg ctgtgatatc agggttcgtg aatgttcact 10620
aggtatgtgc atctgaagtt gaatcacata ggctccagag cagctccatt atattgtcca 10680
cttttattat gctttattcc aagaagtttt ggaagcaact gtgtgacaaa caggataaag 10740

Fig. 12 continu d:

gatgtagaag gctggaggga cagcagtgtt ggctatcatc tgaatctgaa tatttggaaa 10800
atcagctcag actgttacct agggacaggt cactctggtt ggaccataaa actcttcttt 10860
ctttcagttc agtgtaggga attaagctac aaattctgga tggtgcagta ccttagagcc 10920
ctggtgttct ggctgtgtga ggctttcccc atctctcagg agtgggactg ataacgcctg 10980
tggccatcat cctttatcct tcttggtggg tcttttaaaa caaggactgt gtcatgctct 11040
ctacatttaa gaatttcaat ggggccacag gcgcccctag attataaagt tctcataaag 11100
atatgttcat cacatccctt cagctgatct ttgggtttgc agggctaata tttgattccc 11160
aacaatttac cttttcataa tcttgccatt tctacttact gtttagatgg aaaagagtaa 11220
gttctccttg tagactgcag tccacctaac aacttgcctc cagccttgtt aacagtacca 11280
caagcagaac taaaatcatg caaagaattg cagaagcatg gcagtctcag gaagtttcca 11340
ctccaaaaaa accccaaaac aaaaacaaaa acaaaaacaa aaacaaaaaa ccccacccac 11400
tcactaactc acataggtgg tttgtttcct tttgttggac atttacaatt ctatctatct 11460
atctatctat ctatctatct atctatctat ctatctatct atctatctat ctatctatcc 11520
tactatctat ctatctatct atctatctat ctatctatct atctatctat ctatctatct 11580
atctacctac ctacctatct atctatctat ctatctatct atctatctat cctactatct 11640
atctatctat ctatctatct atctatctat ctatctatct atccatccac tgtgtgtgtg 11700
tacatgtcat gcttgtggag gtcagaggac aacctatggg aattgtttct ctctttccac 11760
catgtgtgtt ctggggttca ggtagtcagg ctgggtgata gcatcttccc tgctgggcta 11820
tcccactggc agctttgttg tttcacggaa tctctggaac atcaccctct taatttctct 11880
gttcattggc tgttctttct gtgccttctt tattctgtca tcttcgtctc tcacaatgac 11940
aaatatttaa gttgtcctag gacccagttt tcttttaaat tattttatga ttttatgcac 12000
atgtttcgcc tgcatctatg caccatggtg tgcctcacgc tgctgagttc agaagatatc 12060
aggtcccttg ggacttgact tacagaggtt gtgacaaggc ccagttttct tctgcactgg 12120
cttccttcct gattccatcc taactcttgg cattaaaggc ctgtgtggat tcgatactcc 12180
cagacgtatc tgtgtgctcc cccagaatgt attaaatagc ctcctcaagt tttccatgta 12240
accgtcaaga aggtgtatca aatttaaaat gcccaaagct ggactcttga cacctttata 12300
ccttagcttt tggtattcct cccccatctg gaccagacca tagcttcgtc ttccaaattg 12360
catgtgcaac tttctttttt aatccttagt ttctttcttt tctcttatat tcatatcaca 12420
atgcacccgt tctgtttatt caaccctaag atatacccag cacttaaccg cttcctaccg 12480
tgtatagctt gggcagcgta ggtgaatcca cttttctaac tggattttct gactctatgc 12540
tttatagact ctctctgtat tagtacacag aggggttccc tggaaatata agtcagatga 12600
tggtatgcct ctggaactgt cgatggcttc ctagtttgca tagattaaaa ccaaaggtcc 12660
gggggggggg ggctggagag atggctcagt ggttaagagc agtgactgct cttccaaagg 12720
tcctgagttc aattcccagc aaccatatgg tggctcacaa ccatctgtaa tgagatctga 12780
ctccatcttc tggagtgtct gaagacagct acagtgtact cacatataat aaaaataaat 12840
aaatcttaaa aaaaaaaacc aaaggtcctt ctattggtct agtttgtctt gctccttttt 12900
catcctcgcc catccacatt gctttgacac aggtttccct tgatctcgca ggtttctgct 12960
acatgctctg ccattttgga tgcatctttt gccccctttc cttggatgcc cacatgatgc 13020
catcttacac tgttttgcca ttttatctac ttgttttttt cttttacctt tttgggtttg 13080
ggttaaatgt tgttgtatcc ctgattaatc tttacaaata gctaccaccc ccccccaaga 13140
gtcttattta tttctgatta ctttttttaa tggcacatga ttattttgtt tttgcttctt 13200
taaaaaaacc tttttcattc attgataaaa agcacttact ttccttgatg tgtatgtctc 13260
ctttgtctaa atgtgcaccc cttgtgtgct gtgagtccac aagaaggtgt catatcccct 13320
ggaactggag ttacagatgg ccgtgagcca ccatgtgggt gtggaaatca aaccccagta 13380
cttgggaaga gcagccagtg ctcttagctg ttgataggtc tttccagact gtttttactt 13440
ctgaagatgt gtgtaagtct ctccctctcc acttcttcca cattaagaat ataaggacaa 13500
taaagggaca caatcttgcc tgtgagtttc attgtggtat agttagcaag acgactggat 13560
gaatgaaggc tgcctaagcc ttcggaactt cctggttagc tcatttctta cgtaggaacc 13620
caagaagtct cacgctttac ctctgttgac tgtagagagg ttacttacta ctattttaaa 13680
gtttcaggca agcgtttatg gctgttatat ccccaacctt actcaccgat ccctttcttc 13740
ccatgtgact ttatggaaat tacttaactc cctacatttc agggagctaa tctgaaaaaa 13800
atgtgataaa aatagaatgt agctcaatag aatatagtga agattaagta aagtcatgaa 13860
caaaaagaat ttagttagaa aacctaatgt gtattagggt tctgcaagta cttatgattt 13920
tataactctg ttaataactc attgttttaa gtgcattcat tattcccacc actttgaatc 13980
attattaaga taaatatgat gtttttcaga tcgtccatgc cttaaaaata agcagcatat 14040
agaaaaccta caagaaccag tcctgcaact tttgtttaag ttttcaaaaa tgtaccatcc 14100
agaagaccca cagcatttcg cccacctcat agggaggctt actgaactga gaactctgag 14160
tcacagccac tctgaaatcc ttcgcatgtg gaaaacaaag gaccccaggt tggtgatgtt 14220
attctctgag aaatgggatc tgcactcatt ttcctgaaaa tttacagtgg tgagacttgc 14280
cttaaaattt caaaatcaaa attggatgaa tgatttgaaa aaaatgtgat aactcaaaat 14340

Fig. 12 continued:

```
cattgagtta tcacacacca ccttaaaatg gagatccaaa aggctcatca gggtttgtgg   14400
aactttcaac tcttagaaga tgaacactcc aggattcatc tcaagattgg atgcaaagga   14460
aagaaaaatg gagccatgtg gtggcatagg ctctgtagtt gagagacaga gacaagagga   14520
ccccatgttt aaggcagtcc tgggctatgt agtgagacct catctcagaa gaaaaaagag   14580
aagaactgat ggccatcatt atgtctccaa tcctcaccta aagaaactaa ttttttgttg   14640
ttgttgttta gctgaaacac ataatgtgta aatgtgaaca caggtaggtt atggttagga   14700
taaggcagtt tgtatgcttc actggttaag gacacatgga tggtgtagca gctaatgctt   14760
agaacctctt cttcatgaat gattaacgag gaacaagcct tttcaaattg tatctgcttt   14820
atatttttaa taagtatgtg ataattgtac agaatagtca atattgtaca gaatatttca   14880
atatgggaat ttcatgtgta tatattagat atttctggcg tatttgccac ccaccctgtg   14940
atgcccacac tgccaatgcc cttctcatcc ccctaacatt tcccttcctc tgccttgccc   15000
taaaacacca gattcatacg tgcaagaaaa catttttttaa acaaatttta ttagatattt   15060
tcttcattta caattcaaat gctatcccaa aagtccctta ttccctcccc ttatcctgct   15120
ctccaaccca cccacttctg cttcctggcc ctggtattcc cctgtattgg ggtatataat   15180
ctttgcaaga ccaagggcct ctcctcccaa cgatggtcca ctaggccatc tgcaagaaaa   15240
catttgattc cagtctgagt ctgttttatt tcctttgagt atgactatct cgtcaatatc   15300
tgcaaatgac ataatttcat tcctcttaag gatggaaagg actggattgt gtatgtatat   15360
ctgtatatat tattttattt atccacctgc tgatgacacc tagtttgatt ctgtatcttg   15420
gccactgtga atcgtgctat aataaaccta tgtaggtttt tagattttat gtggctcttg   15480
attctttcag gtgtactccc agaagtggta tggtttatat ggtagctcta tttttagatt   15540
tgaagggaac tatgtagtga ttttcacagt gattaccttc ttctcaatag tgtataaaga   15600
ttaatttgcc ttcaaaattc ttgccaattt ttgtttttcc ctttcctttt ctttttaatt   15660
aaaaaagtgt gtgtgtgtgt gcgtgcatgc caaagaacat tgtggacatc ggaggacagc   15720
tctgtggaat acattctctc cttccacctt tatgggtggt ctcctctgac cccctctgcc   15780
atggcctgca tggcttggtt ctggaagcct gattggggtg atgagcgaat gaagaaatga   15840
cagacatgaa gacacaagta cagaaaagct gagattgggt ggtgtctaag tttctttttt   15900
attgctataa agagacacca ggaccttatg aaagaaagca ttaaattggg ggcttgctta   15960
cagtcttaaa gattgcgtcc atgatgatca tggcaggtag catggcacca ggcatgcact   16020
ggaacagtat ctgaagagct tacattctga tcggcaagtt ggaggcaggg ggtgggggga   16080
gaagaaagtg gctgatgggg cctgggacag gatttggaaa tctcaaagcc catccctagt   16140
gacacacctc ctccaacgag gccacaccca gtcctcctca aataggtcca ccaagtaggg   16200
```

Fig. 13:

```
ccctacttgg tggacctatt tgaggaggac tgggtgtggc ctcgttggag gaggtgtgtc 60
actagggatg ggctttgaga tttccaaatc ctgtcccagg ccccatcagc cactttcttc 120
tccccccacc ccctgcctcc aacttgccga tcagaatgta agctcttcag atactgttcc 180
agtgcatgcc tggtgccatg ctacctgcca tgatcatcat ggacgcaatc tttaagactg 240
taagcaagcc cccaatttaa tgctttcttt cataaggtcc tggtgtctct ttatagcaat 300
aaaaaagaaa cttagacacc acccaatctc agcttttctg tacttgtgtc ttcatgtctg 360
tcatttcttc attcgctcat caccccaatc aggcttccag aaccaagcca tgcaggccat 420
ggcagagggg gtcagaggag accacccata aaggtggaag gagagaatgt attccacaga 480
gctgtcctcc gatgtccaca atgttctttg gcatgcacgc acacacacac acactttttt 540
aattaaaaag aaaaggaaag ggaaaaacaa aaattggcaa gaattttgaa ggcaaattaa 600
tctttataca ctattgagaa gaaggtaatc actgtgaaaa tcactacata gttcccttca 660
aatctaaaaa tagagctacc atataaacca taccacttct gggagtacac ctgaaagaat 720
caagagccac ataaaatcta aaaacctaca taggtttatt atagcacgat tcacagtggc 780
caagatacag aatcaaacta ggtgtcatca gcaggtggat aaataaaata atatatacag 840
atatacatac acaatccagt cctttccatc cttaagagga atgaaattat gtcatttgca 900
gatattgacg agatagtcat actcaaagga aataaaacag actcagactg gaatcaaatg 960
ttttcttgca gatggcctag tggaccatcg ttgggaggag aggcccttgg tcttgcaaag 1020
attatatacc ccaatacagg ggaataccag ggccaggaag cagaagtggg tgggttggag 1080
agcaggataa ggggagggaa taagggactt ttgggatagc atttgaattg taaatgaaga 1140
aaatatctaa taaaatttgt ttaaaaaatg ttttcttgca cgtatgaatc tggtgtttta 1200
gggcaaggca gaggaaggga aatgttaggg ggatgagaag ggcattggca gtgtgggcat 1260
cacagggtgg gtggcaaata cgccagaaat atctaatata tacacatgaa attcccatat 1320
tgaaatattc tgtacaatat tgactattct gtacaattat cacatactta ttaaaaatat 1380
aaagcagata caatttgaaa aggcttgttc ctcgttaatc attcatgaag aagaggttct 1440
aagcattagc tgctacacca tccatgtgtc cttaaccagt gaagcataca aactgcctta 1500
tcctaaccat aacctacctg tgttcacatt tacacattat gtgtttcagc taaacaacaa 1560
caacaaaaaa ttagtttctt taggtgagga ttggagacat aatgatggcc atcagttctt 1620
ctcttttttc ttctgagatg aggtctcact acatagccca ggactgcctt aaacatgggg 1680
tcctcttgtc tctgtctctc aactacagag cctatgccac cacatggctc catttttctt 1740
tcctttgcat ccaatcttga gatgaatcct ggagtgttca tcttctaaga gttgaaagtt 1800
ccacaaaccc tgatgagcct tttggatctc cattttaagg tggtgtgtga taactcaatg 1860
attttgagtt atcacatttt tttcaaatca ttcatccaat tttgattttg aaattttaag 1920
gcaagtctca ccactgtaaa ttttcaggaa aatgagtgca gatcccattt ctcagagaat 1980
aacatcacca acctggggtc ctttgttttc cacatgcgaa ggatttcaga gtggctgtga 2040
ctcagagttc tcagttcagt aagcctccct atgaggtggg cgaaatgctg tgggtcttct 2100
ggatggtaca tttttgaaaa cttaaacaaa agttgcagga ctggttcttg taggttttct 2160
atatgctgct tatttttaag gcatggacga tctgaaaaac atcatattta tcttaataat 2220
gattcaaagt ggtgggaata atgaatgcac ttaaaacaat gagttattaa cagagttata 2280
aaatcataag tacttgcaga accctaatac acattaggtt ttctaactaa attctttttg 2340
ttcatgactt tacttaatct tcactatatt ctattgagct acattctatt tttatcacat 2400
tttttttcaga ttagctccct gaaatgtagg gagttaagta atttccataa agtcacatgg 2460
gaagaaaggg atcggtgagt aaggttgggg atataacagc cataaacgct tgcctgaaac 2520
tttaaaatag tagtaagtaa cctctctaca gtcaacagag gtaaagcgtg agacttcttg 2580
ggttcctacg taagaaatga gctaaccagg aagttccgaa ggcttaggca gccttcattc 2640
atccagtcgt cttgctaact ataccacaat gaaactcaca ggcaagattg tgtcccttta 2700
ttgtccttat attcttaatg tggaagaagt ggagagggag agacttacac acatcttcag 2760
aagtaaaaac agtctggaaa gacctatcaa cagctaagag cactggctgc tcttcccaag 2820
tactggggtt tgatttccac acccacatgg tggctcacgg ccatctgtaa ctccagttcc 2880
aggggatatg acaccttctt gtggactcac agcacacaag ggtgcacat ttagacaaag 2940
gagacataca catcaaggaa agtaagtgct ttttatcaat gaatgaaaaa ggttttttta 3000
aagaagcaaa aacaaaataa tcatgtgcca ttaaaaaaag taatcagaaa taaataagac 3060
tcttgggggg gggtggtagc tatttgtaaa gattaatcag ggatacaaca acatttaacc 3120
caaacccaaa aaggtaaaag aaaaaaacaa gtagataaaa tggcaaaaca gtgtaagatg 3180
gcatcatgtg ggcatccaag gaaagggggc aaaagatgca tcaaaaatgg cagagcatgt 3240
agcagaaacc tgcgagatca aggaaacct gtgtcaaagc aatgtggatg ggcgaggatg 3300
aaaaaggagc aagacaaact agaccaatag aaggaccttt ggtttttttt tttaagattt 3360
atttattttt attatatgtg agtacactgt agctgtcttc agacactcca gaagatggag 3420
tcagatctca ttacagatgg ttgtgagcca ccatatggtt gctgggaatt gaactcagga 3480
```

Fig. 13 continu d:

```
cctttggaag agcagtcact gctcttaacc actgagccat ctctccagcc cccccccccc 3540
ggacctttgg ttttaatcta tgcaaactag gaagccatcg acagttccag aggcatacca 3600
tcatctgact tatatttcca gggaacccct ctgtgtacta atacagagag agtctataaa 3660
gcatagagtc agaaaatcca gttagaaaag tggattcacc tacgctgccc aagctataca 3720
cggtaggaag cggttaagtg ctgggtatat cttagggttg aataaacaga acgggtgcat 3780
tgtgatatga atataagaga aagaaaagaa actaaggatt aaaaaagaaa gttgcacatg 3840
caatttggaa gacgaagcta tggtctggtc cagatggggg aggaatacca aaagctaagg 3900
tataaaggtg tcaagagtcc agctttgggc attttaaatt tgatacacct tcttgacggt 3960
tacatggaaa acttgaggag gctatttaat acattctggg ggagcacaca gatacgtctg 4020
ggagtatcga atccacacag gcctttaatg ccaagagtta ggatggaatc aggaaggaag 4080
ccagtgcaga agaaaactgg gccttgtcac aacctctgta agtcaagtcc caagggacct 4140
gatatcttct gaactcagca gcgtgaggca caccatggtg catagatgca ggcgaaacat 4200
gtgcataaaa tcataaaata atttaaaaga aaactgggtc ctaggacaac ttaaatattt 4260
gtcattgtga gagacgaaga tgacagaata aagaaggcac agaaagaaca gccaatgaac 4320
agagaaatta agagggtgat gttccagaga ttccgtgaaa caacaaagct gccagtggga 4380
tagcccagca gggaagatgc tatcacccag cctgactacc tgaaccccag aacacacatg 4440
gtggaaagag agaaacaatt cccataggtt gtcctctgac ctccacaagc atgacatgta 4500
cacacacaca gtggatggat agatagatag atagatagat agatagatag atagatagat 4560
agatagtagg atagatagat agatagatag atagatagat agataggtag gtaggtagat 4620
agatagatag atagatagat agatagatag atagatagat agatagatag atagatagta 4680
ggatagatag atagatagat agatagatag atagatagat agatagatag atagatagat 4740
agatagatag aattgtaaat gtccaacaaa aggaaacaaa ccacctatgt gagttagtga 4800
gtgggtgggg tttttttgttt ttgtttttgt ttttgttttt gttttggggt ttttttggag 4860
tggaaacttc ctgagactgc catgcttctg caattctttg catgatttta gttctgcttg 4920
tggtactgtt aacaaggctg gaggcaagtt gttaggtgga ctgcagtcta caaggagaac 4980
ttactctttt ccatctaaac agtaagtaga aatggcaaga ttatgaaaag gtaaattgtt 5040
gggaatcaaa tattagccct gcaaacccaa agatcagctg aagggatgtg atgaacatat 5100
ctttatgaga actttataat ctaggggcgc ctgtggcccc attgaaattc ttaaatgtag 5160
agagcatgac acagtccttg ttttaaaaga cccaccaaga aggataaagg atgatggcca 5220
caggcgttat cagtcccact cctgagagat ggggaaagcc tcacacagcc agaacaccag 5280
ggctctaagg tactgcacca tccagaattt gtagcttaat tccctacact gaactgaaag 5340
aaagaagagt tttatggtcc aaccagagtg acctgtccct aggtaacagt ctgagctgat 5400
tttccaaata ttcagattca gatgatagcc aacactgctg tccctccagc cttctacatc 5460
ctttatcctg tttgtcacac agttgcttcc aaaacttctt ggaataaagc ataataaaag 5520
tggacaatat aatggagctg ctctggagcc tatgtgattc aacttcagat gcacatacct 5580
agtgaacatt cacgaaccct gatatcacag caagttagaa tataaaagaa cacaaacatg 5640
gccgaatcgg tgaaggacac aggacaatgg tcccctagga agacgtcgct ggcttgagca 5700
atgggttacc catgctaagt aactgtctca cttcaacctg ggtccggttt gtaacgcata 5760
gtcaagaaga ctgtgaactg acataaagaa gcatgagaaa ttctcctaac acatcagaca 5820
tcttcaccgt gcttcaaaat tccactgaag ggctcacagg aaacaacata tccccttctt 5880
ttctttcaga tgcttatgat actctctaca aaatacacca gtggggcaag taagatgaag 5940
acagagcatt gaccagttga gtcagtgata gcatgtcctg aagatcctga caggttgtag 6000
tggactggag agggtggagt gaaatgtgac tgtggtgaac tcaggagtga gcatgagtaa 6060
ggtggaaact agaatatagc taactccctg agaacctctc tctatcaggg actgggagcg 6120
tatatctgag gggctatgag ggcaagaaag catttcttta agatgcaaga aagatgtaac 6180
ccagtggtaa gcttaaagga atgcacatag agagcaaagc atgatgggat gagagtggta 6240
cattcaactt cagttcatag ttcgataaaa agaacttagt gattgcctgt ccagagatca 6300
ctccgagcct tacttctcag tctgtacctg agaaaagcac tgttgtcgcc gtaagcagag 6360
catattcagt atccgataca tgaagttcac tcattcttct gtagaagtaa gatagtgagg 6420
caataaattc tttagtaata tctgaaaaga ttaatataag tgaattttca tcggaatgga 6480
gcacagaagc aaacacattg tgattgtctt ttcttctctc aaacagtgta agaaaattga 6540
aaagagagtg gggttcaaga tggctaccga tgtgatggcc ggcacagaca atccccttttg 6600
ttgtgccctt gaagcccagg cctgaacctc aagagtgggc gcattaggtg tgctactgaa 6660
aaggaccaga acaaagagag gttgtcatca cgtggaccaa ggaggaggct gtggaactga 6720
gagcctttag cctacagaag ccactctccg tgtatctcta ttatctacag taaggcttac 6780
ccttctaccc ataaaagtgg ttgatactgg gatgttggct gtagattaca tcatcatcct 6840
aagaataatt cagaggcccc aaatggcttc gcaattaaat agggaaagaa ttgccatgat 6900
gaagccaaag taagaaaatg gttacaaata gaattctcac tgatgacctg ccaacactaa 6960
caactgtatt ggctgtgtct ataaatgctt tcaccaccac agagacgcct gtgagctgtg 7020
```

Fig. 13 continu d:

agaaagccca tggccttctc tttgtagtgc aatgaaaagg ggagaaaagg aagcgtgtcc 7080
acccaaggtg atgcttccta ctgcacagtg ttaccagtta gaggagccga agggtagcct 7140
tccttgagaa agttttccgg agaatgaaca ctttcatcag cgctaggatt atgcacctct 7200
agtgtcccag ctgagggctt tgctggtctc atagtaccta gcaaaagaga aaaacacaaa 7260
tgttgatagt ctttcttaat cattatataa atagtaaatt cagtatcttt tcagtatcca 7320
gcataactat tatgtggggg aaaaaaaaac tcctaaaatg atatcattct catttttgtt 7380
actggaacac aaaacagata ttagttcata ttttgtaaaa aaaaaaaaaa gattgtacag 7440
aagaccctcc aaaatgagtt ttgcataatg aaatagaaac taagaatggc tttcttaaac 7500
acacacaccc tccacgaacc actcagtgca ccagaggctg tggagacttt taagcaaaca 7560
ctactcaagc aatcatggac caggtcatca agggcctccc tttagagcct tctggtctag 7620
tgtttgcaga gaacaatcaa taaataagga tattttaaaa ggctgtgcac tgtaacgagt 7680
acagctagtg aaggcagcta gcattgttgg cttttggttt tgattttgtc ctattgcctc 7740
ttcccatgtc ctcaaccaac atgcgacttt acagattgga ctcttgcatt cacacactaa 7800
acaatgatgg ctttcctgct ctattcaaat ggtttattga tttttaaaa aatacttgct 7860
tttttttat tgaagtgagg tcctgcatag ttggttcaac actaacctaa tagggcttca 7920
ccccagtaaa gagacgttcc acaaacagta gcccgtggtt atgagtttca aagcaagcat 7980
gcggccgtcc ccagagctgc tttgtgtctt agaatcacca acgcctttat gtttgatcac 8040
ccataaaagt cctcctcatt ggaagagcct gtctaaggag cacagaagcc ccacccacag 8100
aagccccacc cacagctgag aagctgttgg cagctgctag gggaaggaca atccagtctc 8160
tttagggaac tgagccctct aagttggcca tgcttcagtg gatgacccta cccccacccc 8220
cgtgtatatg agcagcacta tttagagtca atgggttatt ttgaaaaaga aggagcaggg 8280
aaggacatga agttaggaga gaggtgtgtg ggggattcta aaggagctgg tgggaaggaa 8340
gtgaggtgag taagatttaa aaaaaaaaac acttaggaca tgtatacagt tctcacaaga 8400
ataagaaata caatttttaa aagaattaca cggtagagag aaaatgctag atgaatgtga 8460
aactatttat aaacagctac tatgagacat ttgttcctac actaactggt aaaataaaat 8520
gaaaagataa atgaaggct gatgatgtta ggagttagga gttctggctt actgagcaca 8580
gaccctggct tctgtcccca gcaccataga aggcaggcat agaggctcta ggctgatggt 8640
tggcaggtaa gagcagacac agcagggctt taagatcctc cttagctaca tactgagttc 8700
tacatagcct ggactacatg agagtctatt tccaaaaaac agtgtttaaa agtgtgtgcg 8760
tgcgtgcgtg cgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtc tgtctgtctg 8820
tctgtctgtc tgtgtgtact tgaagttgaa taaaattctg tttttctttt tacttccaga 8880
ggttgagtct ttcccaccat aaagctgggc tacatgaagg aacatcactt cagtttttga 8940
cgccttctgt aatgcagcct gatcctcagt ggttaaattt tcaaatcctt aaaaataaat 9000
aaataaataa ataaataaag cataaagaaa cagataatca aaagagaacg ttcataaatt 9060
aaatgtttaa aaatatttac ttattaaccc taagtgcttc tggggaataa caatagtata 9120
taaattgaaa tctgaataaa attctttttt tctattttgg ggcactttta tcagcactag 9180
gttacacttt cttacagcaa ggacaaatca acatagctgc attagccttg aagaatgtaa 9240
ttttgtaaag agtttctgta cctctaagca tataataaag ctgagttata tataagtttt 9300
ggtcaaactg tgacttgttt agtattatag aggaaacctt ctttgtttct tgtttcatct 9360
aagaaagccc acatcattat aataattttt gtaaatattt agattgtcac tcatctgatt 9420
aattgacaaa gacaaataaa catttttttt tttttagtt agaagcaaca ttgctagaaa 9480
acccagggct tcacacacat tagacaggtg ctctgcctct gagcgataga tctagtacca 9540
caactagaaa tttctgtttg atatactatc cctttaaaaa aaagaaaatt tggggctgga 9600
gagatagctc agtgggtaag agcactgact actcttccta aaaagccctg agttcaagtc 9660
ccagcaacca cacggtggct cgcaaccatc cgtaacgaga tctaacgccc tcttctggtg 9720
cacctgaaga cagctacagt gtacttatga aatctttggg ccagaacgag cggggccgac 9780
cagagtgagc agaggtccta aaaattcaat tctcaacaac cacatgaagg ctcacaacca 9840
tctgtacagc tacaatgtac tcagataaat aaataaataa ataaataaat aaataaataa 9900
ataaataatc tttttttaaaa attgggggct agaaagacag atcagcagtt aagagcctgt 9960
cttgctcttg cagagactct ggttccaacg cccacatggt aattcacagc catcggtagg 10020
aaactctagt tctaggggac caaacactct cttctggcct ccacaatatg ttgtgtacat 10080
gatatgaata tataattgca atcaaaacat gcataacaca taaacataaa aataaataaa 10140
tccttaaaaa ataatgaaac agaattttac ctgggagtcc cttggtaaac ttcattagcc 10200
cttgtatgtg caggactgat acctctgaga gtctcagaaa acttagttcg gggttggaac 10260
cctcctgcag ctggtccata aaaaaaataa ataaataaaa atgattcata gtatgtcttt 10320
ctgataatta ctaaggtgtt tcttttttac tgttcttgct atcatttcct aatatcattc 10380
attacaaatt atttcttgcc agcaatggat acatacaagt ttgcttgttt ctcccaaggg 10440
aatcatggat ttttggtgag cagtcactat ggtattcaga agccgatgtt cctcttgagt 10500
tagagtcatg ttgtcctgaa cctataagga aacagagggc gcctgtaatg tgtgcttgtt 10560
acacactgtt cataaagaca caagagactg tatggaagaa gcgtcttgcc cctttcccag 10620

Fig. 13 continued:

atctggtgct ggatgacacg tgtttggtgt ctgctccttc atcttccact tggatggcag 10680
ggtacagggc aggcccgtgc ttgaagttct tgcgaagtct ctttgactta cactggatct 10740
ctgtgagcaa acctgtgaca ttacagaatt cacattaagt actttggct cggaaaacaa 10800
cactgtaggg tttgattaaa atggtcttca gtgagtaaat agtcctagat atctctcgtg 10860
catgtctgca tctattaaat ttcctattct gcttttccct tcatagacac cttgtctgct 10920
taaaaatagt tataacaaag attaagaatt aatttttcat ttccaaatta tgtttagatt 10980
tttacataag agatttacac attagttgca tcttggaagt gaaaaagtca tttgatagaa 11040
tagtagaatt acattatcta tgactattat cactggtttt attaataagg ttaaaaaaga 11100
ctatcttagg atttttaaaa ataagctaag tcaatttttc ctatgtcaga caaagcattt 11160
aattcttttt caaaggcaaa taatatttat tttaggattg tttttgtttg tttttaaata 11220
ttttataatt acttgcatga gtctgtatgg ggaaatgcac aggtgagtgc aggtgcccac 11280
agagcacaga ggggtcatag ccccttgaag tgagactata gctggttgtg agctgtagtg 11340
tatgtactgg actggactgg actactctaa gtacttttag cagctactaa aagtatccaa 11400
ctattaatat acctgtcgtt ttcaaagtat ttttaaaaac acacctataa aataaaaaac 11460
taatgaggag ctgaagtata aacaaaaacc ttaaaacccc tctatattat tttattattt 11520
gtcagagaac agaatacaca aaaaaataaa gccattttaa gcatattgag tattgaacat 11580
ggtagtaaac acctataacc ccagctactt aagacactgt tacaagatga tctcacgttg 11640
aataccagcc atggcaactc agtgtttttt agtaggaaag gagtctatgg gtgtagctta 11700
gtggtagatt gtctaggatg tgagaaacct gaactcaacc cccaggacca tgagaaggaa 11760
aaccacaaga agagaacgaa gaggaggaaa aggaggaaga agagggggaa gaggaggaag 11820
aggaaaagaa aggaggaaag ggaaaggaga atgagggaag gaagaggagg aggggccgtc 11880
agagcactaa cacatggcag tattaggaca actgaaaggt aaatgaagtg tgcagatcca 11940
agcaatggtg gcagggtgcg gaccagctta tgctcgccaa tcaccgtgac tgctcaccaa 12000
ctctctcgag cttcaccttg ccagaatttc tttaagtttc tttaagtgta ggatgactca 12060
ttttatgcat agagttcctc agaggactcc aacttccctt tcacatgcat tgcctgccac 12120
aggttaaaag acaacttttg agccagaaat tggagggcag aggcaggggg atctctgagt 12180
ttgatgccag cctggcgtac atagcaagtt ccaggtcagc tacatagtga gatcctgaat 12240
acaaattaaa ttaatgaata aaataaaaga caacttttc aaaggtttct ttggactgag 12300
ttgtatctcc cactgctcat atgctaatgt gttgatattc agagatgagt cctgacactc 12360
~~gtgagactat tagtggactt tacatatgag gatgtcaggg gtccagagat ccagttaccc~~ ~~12420~~
ttcaattgtc ctaagactgt agtgtgttag gactctgaga acccttctc agaaagggaa 12480
gacttccttt tctcttcttt ctcctcctct ttctttcttt ttgggaacta ctaactttaa 12540
ctttgcactt ttcaaggcag tatatatttt taggttaagt aacagaaaca tacaatacct 12600
tttttaaatc actatggagg tttcttttat tttatgtgtg tgttttgctt gcatgtgtgt 12660
atgcacacaa cactcatacc tgcagaggtc agaagagggc ctgggatgct caggaactga 12720
gttataaatg attgtgaggg gccacgtgga agctggggac cgggcccagg tcctctccta 12780
atgctcttaa ctgcctaact actcttcagg tctcagagaa catctaatta agatgaaact 12840
atgactgatt tcaaaattag agcaatatag aacagggcaa catttatggg aatcactact 12900
taaattgata atgggccatc cacgtggcgc aattactgta catatgtgtg tggtgtgcat 12960
atacaaacga atacggccat gttcaaaata actcttgaag aatgctttat gtcctcaaga 13020
aggttggaaa tttttaattt tttattaatt tttaaattca tttttattat atagaggaga 13080
gtgtgaatac agtccttcac taacacaaat gatgcacctg agtttcccac atttgtggaa 13140
atcacagggg tcagtcagca catccctagt gcaatgggcg attcttgccc tggaaaaacc 13200
accaccacta tcatgtcctc cttgtcagga aggtaaggct gggaaaagtt aaatcccctc 13260
agaaaagcaa gccacgaaca cactggcact tacattctgc caacatcccc accgccttac 13320
acttcttcag tctgcactct tggcatttc tgcgcatgta catgtccatt tcacagtgac 13380
caccgttctt gcaagaatac actgcattct tggtaatgct acgccggaaa aagcctaaca 13440
acaaagatgg tcaaagtttg acatggatat gtttagttac ttctcatttt tattaacatt 13500
ttacctttat agaaagtaat caagctaact agagtatttc aattaaatag ggtttttttt 13560
attcccgaca tttttttaaa tatacattta tctggaaagt tagctaattc tataattctt 13620
tatagttagt atagttcagt tgacatcctg aggctgggtc tttctttata acgtctggag 13680
attaaatgga gtatcattat cacattaaga caactctcta ccctactata ggtaaattac 13740
ctataagaat ttttttttta actgcattat acaaaatatc aagatcctac aatgcatctc 13800
tcgagtttcc tttgagagaa tacatggtca cctaaaataa agaaaaaatt cagaaaagcc 13860
tctcttttct aaaccctccc agtctgagtt gcagcccaca aaggtggtcc ctcccctccct 13920
catccgcaga tgctctctct ccctttccca gcacttgcca gatagttccc atcccttttc 13980
ctccttcgaa ccttattaca cagtaatacc tgccatttcc tgctctccct cctatgagtt 14040
gtcccaaagc cccaccactt actctcttgc agtggacact tgaaagatgg agtttcgcca 14100

Fig. 13 continu d:

```
tttcagttca acgaaggaag aaaaagctca gctcaggttt tcaaaactgt gctgggttct   14160
tcagctcagt tttaacagta acccgtcacc gtcaagagca ggccacccct gcctaactac   14220
tcgccacagg actctgtcac ctcatatcct ttggctattc tttgccagtg ttgaggttta   14280
agatttacag gtaaaatttt accctaaaca aacacgttta tttgtgttac attgactaaa   14340
ctgttttctg ttgtttgggt tgttttgttt tgtttcagag ggtgctagaa tcaaatccaa   14400
aagtaggtgc taagtgctct accactggcc tgcattccca tcccgatgtt tatccatttc   14460
tggagtggta gactgtttca gaactgttat ttatcagaga ggcttccgct ctttacaaat   14520
atcattgcaa acactggccc ttgcttcccg actttcagct tgttctcttt ctatacgaaa   14580
cactttttaa aatttattga cttagaatca tagaagtcat ttatttcaaa ttttctccaa   14640
gtgtataaat tttatcaacc acataaaaag gtgtcatggt ttacagatct ttacattgtc   14700
atgttttgtt ccagcatcaa attcttctag aaatagggaa aacgtttgct tatatgtatc   14760
gagtttaaag taaagtgtga aagtcttcat caaaaatgaa gagacagacc tggtagtgaa   14820
tgctcttaat gccagcagta aggatacaaa ggtaggtgga tctctgagag ttcaaggaca   14880
gcctggtcta aacagtgagt tccaggacag ccagggatgc atagagagac cttttctcaa   14940
aaaaagaaga aagatgagga ggagaagatg aggacgaagg ggaagacaag ggaggagaaa   15000
aaaaaataaa gaaagaaatg gtccaggcat ggtggcccac acctttaatc ccagcaccta   15060
gaaagcagag acaggcaggt atctgtgagt tcaaagccaa cctgatctcc atagtgaggt   15120
ctaggccagc cagggctata caatgatact ctacttcaaa agagagagag gggggagggg   15180
agggagagag agagagagag agagagagag agagagagag agagagagag agagagagag   15240
agagagagag agagagggag agagagagag agagagagag agagagagag agagagagag   15300
agagagagag agaaagaaag aaagaaagaa agaaagaaag aaagaaagaa agaaagaaag   15360
aaagaaagaa agaaagaaag aggaagggaa agggaaaggg aaagggaaag ggaaaggaaa   15420
ggaaaggaaa ggaaaggaaa ggaaaggaaa ggaaagaaac taaccttaaa cttaacattt   15480
tattctccta tctttgatca cacagattga tccctttaga catacttttt cctggttttg   15540
tttctcatgt tttataacga ttgcacacgc acgtgcacac acacatacac acgtgtgcgc   15600
gtgcgcagca ctgtctgttt ggttgcgcct tacccttacc tttgcacccc tcacaagtaa   15660
gtgcattgta gtgatatcct gatgccttat caccgcagac catgcagagc atgtcatctt   15720
gacccttcca gcgtaaagaa caagttgctc tgcacctttg tgccccaggg aattcatctt   15780
catcatcgtg aacaaccaca taagtggact gaccagattc tcgagcgtcg agtccacacc   15840
aactgcctcc actgtactgt ggatccaagc catgtgtgtt gaaatgactt tgtaaagatg   15900
gggactgtaa cgctggagga aactgagcgg tagaatattg acaatagggt ggttcttgga   15960
aatctgtatc acacagctga taatggaatt gttctggcaa aatatctggt tttattaaca   16020
taaaaaagaa caaatcataa aaggtcaaaa aagttcataa ttgttttaaa cctcaaaatg   16080
ttttatactc aacagaaagt tagttcaaca catctattcc atctcaggac ttgccgtcca   16140
gcacaatagg aacacagtaa atggaactaa caatacctaa taaatgctat gaccccaaag   16200
```

Fig. 14:

cDNA splice variant of L66 from Mus musculus (SEQ ID NO. 22)

```
atgttaataa aaccagatat tttgccagaa caattccatt atcagctgtg tgatacagat   60
ttccaagaac caccctattg tcaatattct accgctcagt ttcctccagc gttacagtcc  120
ccatctttac aaagtcattt caacacacat ggcttggatc cacagtacag tggaggcagt  180
tggtgtggac tcgacgctcg agaatctggt cagtccactt atgtggttgt tcacgatgat  240
gaagatgaat tccctggggc acaaaggtgc agagcaactt gttctttacg ctggaagggt  300
caagatgaca tgctctgcat ggtctgcggt gataaggcat caggatatca ctacaatgca  360
cttacttgtg aggggtgcaa aggctttttc cggcgtagca ttaccaagaa tgcagtgtat  420
tcttgcaaga acggtggtca ctgtgaaatg gacatgtaca tgcgcagaaa atgccaagag  480
tgcagactga agaagtgtaa ggcggtgggg atgttggcag aatgtttgct cacagagatc  540
cagtgtaagt caaagagact tcgcaagaac ttcaagcacg ggcctgccct gtaccctgcc  600
atccaagtgg aagatgaagg agcagacacc aaacacgtgt catccagcac cagatctggg  660
aaaggggttc aggacaacat gactctaact caagaggaac atcggcttct gaataccata  720
gtgactgctc accaaaaatc catgattccc ttgggagaaa caagcaaact tctgcaggag  780
ggttccaacc ccgaactaag ttttctgaga ctctcagagg tatcagtcct gcacatacaa  840
gggctaatga agtttaccaa gggactccca ggatttgaaa atttaaccac tgaggatcag  900
gctgcattac agaaggcgtc aaaaactgaa gtgatgttcc ttcatgtagc ccagctttat  960
ggtgggaaag actcaacctc tggaagtact atgagaccag caaagccctc agctgggaca 1020
ctagaggtgc ataatcctag cgctgatgaa agtgttcatt ctccggaaaa ctttctcaag 1080
gaaggctacc cttcggctcc tctaactgga agaatgagtg aacttcatgt atcggatact 1140
gaatatgctc tgcttacggc gacaacagtg cttttctcag atcgtccatg ccttaaaaat 1200
aagcagcata tagaaaacct acaagaacca gtcctgcaac ttttgtttaa gttttcaaaa 1260
atgtaccatc cagaagaccc acagcatttc gcccacctca tagggaggct tactgaactg 1320
agaactctga gtcacagcca ctctgaaatc cttcgcatgt ggaaaacaaa ggaccccagg 1380
ttggtgatgt tattctctga gaaatgggat ctgcactcat tttcctga             1428
```

cDNA splice variant of L66 from Mus musculus (SEQ ID NO. 23)

```
tcaggaaaat gagtgcagat cccatttctc agagaataac atcaccaacc tggggtcctt   60
tgttttccac atgcgaagga tttcagagtg gctgtgactc agagttctca gttcagtaag  120
cctccctatg aggtgggcga aatgctgtgg gtcttctgga tggtacattt ttgaaaactt  180
aaacaaaagt tgcaggactg gttcttgtag gttttctata tgctgcttat ttttaaggca  240
tggacgatct gagaaaagca ctgttgtcgc cgtaagcaga gcatattcag tatccgatac  300
atgaagttca ctcattcttc cagttagagg agccgaaggg tagccttcct tgagaaagtt  360
ttccggagaa tgaacacttt catcagcgct aggattatgc acctctagtg tcccagctga  420
gggctttgct ggtctcatag tacttccaga ggttgagtct ttcccaccat aaagctgggc  480
tacatgaagg aacatcactt cagtttttga cgccttctgt aatgcagcct gatcctcagt  540
ggttaaattt tcaaatcctg ggagtccctt ggtaaacttc attagccctt gtatgtgcag  600
gactgatacc tctgagagtc tcagaaaact tagttcgggg ttggaaccct cctgcagaag  660
tttgcttgtt tctcccaagg gaatcatgga tttttggtga gcagtcacta tggtattcag  720
aagccgatgt tcctcttgag ttagagtcat gttgtcctga acccctttcc cagatctggt  780
gctggatgac acgtgtttgg tgtctgctcc ttcatcttcc acttggatgg cagggtacag  840
ggcaggcccg tgcttgaagt tcttgcgaag tctctttgac ttacactgga tctctgtgag  900
caaacattct gccaacatcc ccaccgcctt acacttcttc agtctgcact cttggcattt  960
tctgcgcatg tacatgtcca tttcacagtg accaccgttc ttgcaagaat acactgcatt 1020
cttggtaatg ctacgccgga aaaagccttt gcacccctca caagtaagtg cattgtagtg 1080
atatcctgat gccttatcac cgcagaccat gcagagcatg tcatcttgac ccttccagcg 1140
taaagaacaa gttgctctgc acctttgtgc cccagggaat tcatcttcat catcgtgaac 1200
aaccacataa gtggactgac cagattctcg agcgtcgagt ccacaccaac tgcctccact 1260
gtactgtgga tccaagccat gtgtgttgaa atgactttgt aaagatgggg actgtaacgc 1320
tggaggaaac tgagcggtag aatattgaca atagggtggt tcttggaaat ctgtatcaca 1380
cagctgataa tggaattgtt ctggcaaaat atctggtttt attaacat             1428
```

Fig. 14 - continu d:

Protein splice variant of L66 from Mus musculus (SEQ ID NO. 24)

```
MLIKPDILPE QFHYQLCDTD FQEPPYCQYS TAQFPPALQS PSLQSHFNTH GLDPQYSGGS    60
WCGLDARESG QSTYVVVHDD EDEFPGAQRC RATCSLRWKG QDDMLCMVCG DKASGYHYNA   120
LTCEGCKGFF RRSITKNAVY SCKNGGHCEM DMYMRRKCQE CRLKKCKAVG MLAECLLTEI   180
QCKSKRLRKN FKHGPALYPA IQVEDEGADT KHVSSSTRSG KGVQDNMTLT QEEHRLLNTI   240
VTAHQKSMIP LGETSKLLQE GSNPELSFLR LSEVSVLHIQ GLMKFTKGLP GFENLTTEDQ   300
AALQKASKTE VMFLHVAQLY GGKDSTSGST MRPAKPSAGT LEVHNPSADE SVHSPENFLK   360
EGYPSAPLTG RMSELHVSDT EYALLTATTV LFSDRPCLKN KQHIENLQEP VLQLLFKFSK   420
MYHPEDPQHF AHLIGRLTEL RTLSHSHSEI LRMWKTKDPR LVMLFSEKWD LHSFS         475
```

Fig. 15

Human:

SEQ ID NOs: 1 = all exons human	Fig.1
SEQ ID NOs: 4 = NC fragment	Fig.5
SEQ ID NOs: 6 = poly A1 fragment	Fig.6
SEQ ID NOs: 8 = poly A2 fragment	Fig.7
SEQ ID NOs: 10 = GF fragment	Fig.8
SEQ ID NOs: 12 = SF fragment	Fig.9
SEQ ID NO. 3 = protein L66 from SEQ ID NO. 1	Fig.2
SEQ ID NOs: 2 = RC of all exons human	Fig.1
SEQ ID NOs: 5 = RC NC fragment	Fig.5
SEQ ID NOs: 7 = RC poly A1 fragment	Fig.6
SEQ ID NOs: 9 = RC poly A2 fragment	Fig.7
SEQ ID NOs: 11 = RC GF fragment	Fig.8
SEQ ID NOs: 13 = RC SF fragment	Fig.9
SEQ ID NOs: 14 = primer 18 sRNA fwd	No Fig.
SEQ ID NOs: 15 = primer 18 sRNA rev	No. Fig.
SEQ ID NOs: 16 = probe 18 sRNA	No. Fig.

Mouse:

SEQ ID NOs: 17 = cDNA L66 from Mus musculus	Fig. 4
SEQ ID NOs: 18 = RC of cDNA L66 from Mus musculus	Fig. 4
SEQ ID NOs: 19 = protein L66 from Mus musculus	Fig. 4
SEQ ID NOs: 20 = genomic DNA L66 from Mus musculus	Fig. 12
SEQ ID NOs: 21 = RC of genomic DNA L66 from Mus musculus	Fig. 13
SEQ ID NO.: 22 = cDNA L66 from Mus musculus (splice variant 1)	Fig. 14
SEQ ID NO.: 23 = RC of cDNA L66 from Mus musculus (splice variant 1)	Fig. 14
SEQ ID NO.: 24 = protein L66 from Mus musculus (splice variant 1)	Fig. 14

Fig. 16

```
QUERY     1  PGIKRSRPTYSSSRNKGQEEFCVVCGDKASPSPYHYNALTCEGCKEIPMVKNFKTFLLGF
             P .K:.R   S:.R KG :E.CVVCGD:AS   YHYNALTCEGCK             GF
HIT     107  PVTKKPRMGASAGRIKG-DELCVVCGDRASG--YHYNALTCEGCK-------------GF

             FQCSIXQNAVYSCRNGSHCEMDMYMRRKCQECRLKKYKAVGMLAEC----LLTEIQCKLK
             F: SI.:NAVY.C:NG.:C MDMYMRRKCQECRL:K K.:GMLAEC    LLTEIQCK K
             FRRSITKNAVYKCKNGGNCVMDMYMRRKCQECRLRKCKEMGMLAECMYTGLLTEIQCKSK

             RLQKNFKEKNHFYSNIKVEEEGVDHSFLSSTTRPGKESMELTEEEHQLINNIVAAHQKYT
             RL:KN.K:  H  ..:. :.EG D . ::STT:. :E..ELT.::..L:: I: ::.K..
             RLRKNVKQ--HADQTVNEDSEGRDLRQVTSTTKSCREKTELTPDQQTLLHFIMDSYNKQR

             IPLEETNLYLQEHTNPELSFLQLSETAVLHIRGLMNFTKGLPGFENLANEDQTALQKGSK
             :P E TN .L:E. :.E :FL L:E.A. H:: L:.FTK LPGF:.L :EDQ.AL KGS.
             MPQEITNKILKEEFSAEENFLILTEMATNHVQVLVEFTKKLPGFQTLDHEDQIALLKGSA

             TEVIFLHGAQLYNTMIISICLILPYVWMKIHFRISFLGVTEEFITXLFYFYKRMSKLDVT
             .E.:FL..A::N..: S    L      :. RI.  G:::E:IT :F FYK.:.:L.:T
             VEAMFLRSAEIFNKKLPSGHSDL------LEERIRNSGISDEYITPMFSFYKSIGELKMT

             NTEYALLAATIVFS-DRPCLKNKQYMENLXEPVLQILYKYSKMYHPEDPXHFAHLIWKHT
             ..EYALL.A.::.S DR. :K::: :E.L.EP:L.:L.K..K::.PE:P.HFA L: : T
             QEEYALLTAIVILSPDRQYIKDREAVEKLQEPLLDVLQKLCKIHQPENPQHFACLLGRLT

             ELRTLNYNHSEILSTWKTKDPKLATLLSE      384
             ELRT.N::H:E:L.:W:..D K...LL.E
             ELRTFNHHHAEMLMSWRVNDHKFTPLLCE      471
```

Fig. 17 A

#	Exon		Intron	
	Start	Stop	Start	Stop
1	ATG-> 180	561	562	2766
2	2767	2907	2908	5448
3	5449	5591	5592	5679
4	5680	5784	5785	5930
5	5931	6030	6031	7213
6	7214	7327	7328	8963
7	8964	9086	9087	9759
8	9760	9873	9874	1409
9	14010	14257(incl. TGA)		

Fig. 17 B

#	Exon		Intron	
	Start	Stop	Start	Stop
1	ATG-> 180	561	562	2766
2	2767	2907	2908	5448
3	5449	5591	5592	5679
4	5680	5784	5785	5930
5	5931	6030	6031	7213
6	7214	7327	7328	8963
7	8964	9086	9087	9801
8	9802	9873	9874	1409
9	14010	14257(incl. TGA)		

Fig. 18 A

Exon	Intron	Exon
... GGGTGCAAAG	GTAAGGGTAA>>>>>TTGTTGTTAG	GCTTTTTCCG...
...TTGGCAGAAT	GTAAGTGCCA>>>>>AATGTCACAG	GTTTGCTCAC...
...TGGGAAAGGG	GCAAGACGCT>>>>>TTCCTTATAG	GTTCAGGACA...
...AAGCAAACTT	GTATGTATCC>>>>>TATGGACCAG	CTGCAGGAGG...
...GGACTCCCAG	GTAAAATTCT>>>>>TATTTTTAAG	GATTTGAAAA...
...ACCTCTGGAA	GTAAAAAGAA>>>>>CTTTTGCTAG	GTACTATGAG...
...CCTCTAACTG	GTAACACTGT>>>>>ATCTTTTCAG	ATATTACTAA...
...CTTTTCTCAG	GTACAGACTG>>>>> TGTTTTTCAG	ATCGTCCATG...
...ATTTTCCTGA		

Fig. 18 B

Exon	Intron	Exon
... GGGTGCAAAG	GTAAGGGTAA>>>>>TTGTTGTTAG	GCTTTTTCCG...
...TTGGCAGAAT	GTAAGTGCCA>>>>>AATGTCACAG	GTTTGCTCAC...
...TGGGAAAGGG	GCAAGACGCT>>>>>TTCCTTATAG	GTTCAGGACA...
...AAGCAAACTT	GTATGTATCC>>>>>TATGGACCAG	CTGCAGGAGG...
...GGACTCCCAG	GTAAAATTCT>>>>>TATTTTTAAG	GATTTGAAAA...
...ACCTCTGGAA	GTAAAAAGAA>>>>>CTTTTGCTAG	GTACTATGAG...
...CCTCTAACTG	GTAACACTGT>>>>> TACTTCTACA	GAAGAATGAG...
...CTTTTCTCAG	GTACAGACTG>>>>> TGTTTTTCAG	ATCGTCCATG...
...ATTTTCCTGA		

Fig. 19:

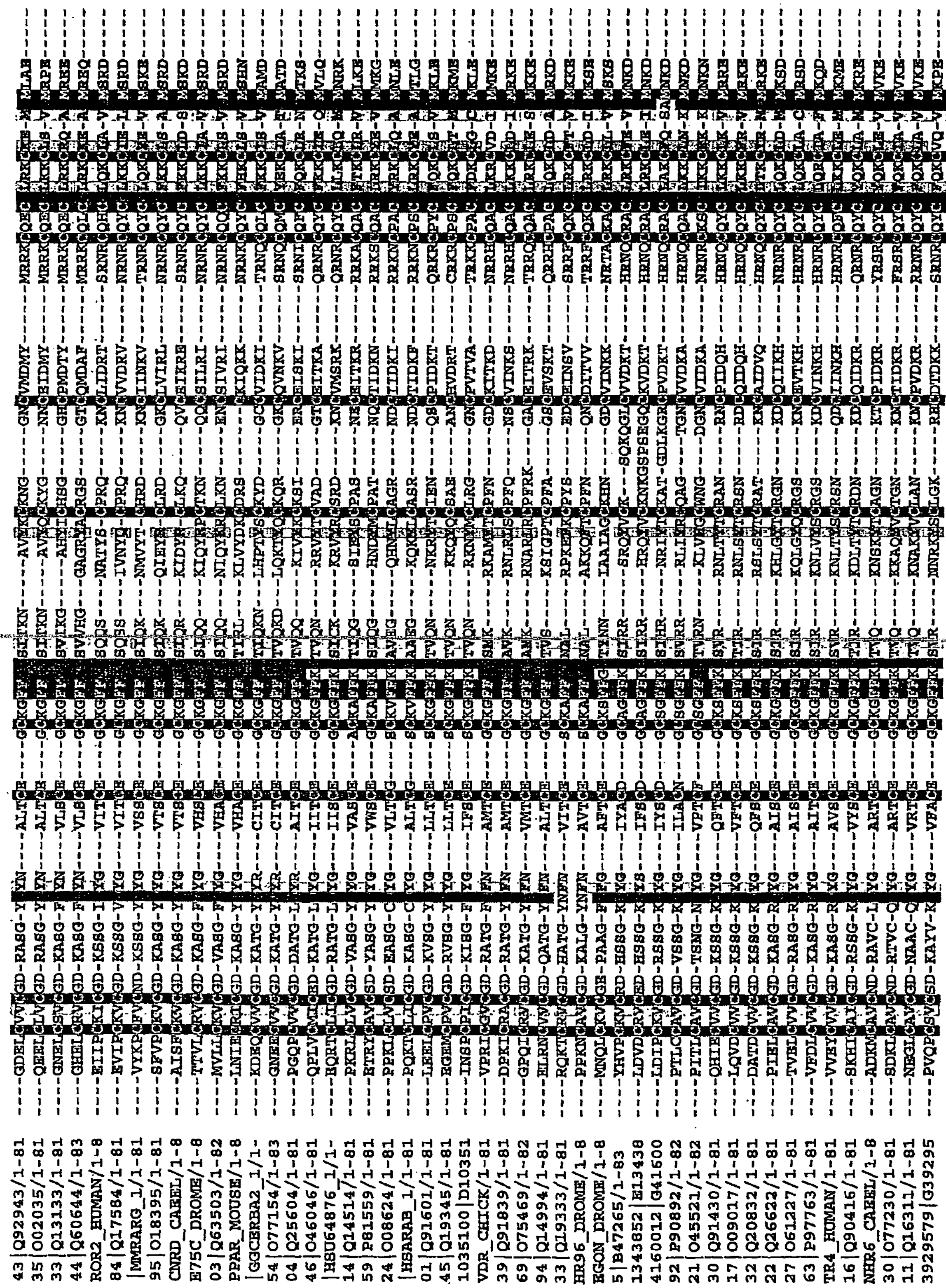

Fig. 19 continued:

```
51|O02151/1-81    ----SADFCVVCGD-KAIG-K[illegible]YG----AVACN----GCKGFF[illegible]SVW-----QNLQYTCRFN-----KQCNIDKD------HRNACRYC[illegible]FQKCLA-D[illegible]MKPE-----
66|O45666/1-81    ----LAEPCAVCGD-KSTG-T[illegible]YG----VISCN----GCKGFF[illegible]TVL-----RDQKFTCRFN-----KRCVIDKN------FRCACRYC[illegible]FQKCVQ-V[illegible]MKRE-----
55|Q92655/1-81    ----VSALCAICGD-RATG-K[illegible]YG----ASSCD----GCKGFF[illegible]SVR-----KNHMYSCRFS-----RQCVVDKD------KRNQCRYC[illegible]LKKCLR-A[illegible]MKKE-----
NH64_CAEEL/1-8    ----ENVFCAICGD-RATG-K[illegible]YG----AMSCD----GCKGFF[illegible]TIR-----KRHSMVCRFG-----EKCQVDKA------KRNSCRKC[illegible]EDVCLR-K[illegible]MRRD-----
29|P91829/1-81    ----VEEICHICND-KSTG-K[illegible]YG----AISCD----GCKGFF[illegible]SIR-----KRYHYQCRFE-----QNCDVTKN------KRNACRAC[illegible]LQKCVK-A[illegible]MKSN-----
1347627|E13476    ----ESTICSVCCD-EASG-R[illegible]YG----VVACF----GCKGFF[illegible]TVR-----AGKNYVCRYS-----KKCRIDKA------GRNVCRSC[illegible]FQKCLE-V[illegible]MEPD-----
NH35_CAEEL/1-8    ----DNSICHICSD-VATG-R[illegible]YG----AIACN----GCKGFF[illegible]TVR-----RNYEYHCRFE-----SKCEIDKH------NRAVCRYC[illegible]FMKCLS-S[illegible]MRDD-----
1347154|E13471    ----TNRICAVCGD-TPAK-I[illegible]YG----VLACF----GCKGFF[illegible]AVKD-----GRNKYVCRFE-----KNCEVTKF------ERNACRYC[illegible]FRKCLL-V[illegible]MNPD-----
NH21_CAEEL/1-8    ----GDSVCAVCGD-GIAK-L[illegible]YG----VLACY----GCKGFF[illegible]TLT-----GKYRHACRFS-----NNCIVDKF------QRNSCRYC[illegible]FQRCLQ-A[illegible]MDPK-----
1347738|E13477    ----HCRKCVVCGD-RACSHLYYG----VAACH----GCKCFFW[illegible]TVK-----SRLNYVCRYG-----GNCSISTA------GRNACRYC[illegible]FHRCLF-V[illegible]MKIE-----
3253108|G32531    ----KKDRCMVCGD-NSTG-Y[illegible]YG----VQSCE----GCKGFF[illegible]SVHK-----NIAYVCTKG-----ENCTFSYEN(6)--VRTRCQAC[illegible]EAKCLA-V[illegible]MNRD-----
50|Q18150/1-81    ----DKHHCSVCGD-RPTG-Y[illegible]YD----VLSCN----GCKTFF[illegible]TIIN-----SRNEICTKG-----GNCQFTKD------FRCACRAC[illegible]FEKCVR-V[illegible]MNPS-----
NH10_CAEEL/1-8    ----PEEVCLVCSD-ISTG-Y[illegible]YG----VPSCN----GCKTFF[illegible]TIMK-----NQTFSCQFQ-----GKCPVDKS------IRCACRHC[illegible]FEKCLQ-V[illegible]MDRN-----
1347565|E13475    ----NKEVCLVCQD-FSSG-Y[illegible]YG----IPSCN----GCKTFF[illegible]TVMK-----KQKFVCQFD-----QNCPVDKS------IRCACRFC[illegible]FEKCLK-V[illegible]MDKT-----
78|Q21878/1-81    ----EGELCAVCSD-LATG-Y[illegible]YG----VASCN----GCKTFF[illegible]TIVS-----EQTFICQYN-----GNCDVNKN------IRCACRHC[illegible]FNKCLL-V[illegible]MDAK-----
CSR1_CAEEL/1-8    ----PGTLCAVCDD-IATG-K[illegible]YS----VASCN----GCKTFF[illegible]ALVN-----NREFVCQGN-----KDCPVNKG------VRCACRYC[illegible]LQKCLA-V[illegible]MDKN-----
89|Q17589/1-81    ----PGELCSVCGD-VASG-I[illegible]YS----VAACN----GCKTFF[illegible]VVLE-----NRTYSCKNN-----GDCIIDKS------MRCSCRHC[illegible]YKKCLI-A[illegible]MDRG-----
72|Q21372/1-81    ----HTEQCVVCGD-AADG-F[illegible]YG----VRSCR----GCNAFF[illegible]AVTF-----NMSFTCRRG-----GRCPVDKN------ARCACRAC[illegible]EAKCLA-V[illegible]MDKK-----
4139072|G41390    ----GRLICDVCG--DVAFGK[illegible]YG----INACN----GCKGFF[illegible]SVWS-----RRQYSCRFG-----GDCPVVKE------HRNVCRSC[illegible]LKKCFE-V[illegible]MNPD-----
1348022|E13480    ----INLVCVVCG--DQAFGK[illegible]YG----VNACN----GCKGFF[illegible]SVWH-----NRQYLCRFE-----GRCAIAKE------HRNVCRAC[illegible]LKQCFV-A[illegible]MNPR-----
89|O62389/1-81    ----KSLQCAVCG--DVALGK[illegible]YG----VNACN----GCKGFF[illegible]SIWK-----NRTYACRHG-----GKCLVAKE------QRNACRSC[illegible]LTRCLD-V[illegible]MNPR-----
11|O17611/1-81    ----SDTSCLVCG--DPHGKR[illegible]YG----AMSCN----GCKGFF[illegible]SIWE-----KRTYKCSFN-----NECLIEFK------YRNRCRAC[illegible]LKRCLH-V[illegible]MDAN-----
90|O16890/1-81    ----LVSACLVCG--DPNIQR[illegible]YG----TVSCN----GCKGFF[illegible]SIWE-----KRTYKCSFH-----NQCSIEHK------YRNRCRAC[illegible]LEKCLR-A[illegible]MDAS-----
06|Q21006/1-81    ----TGLKCRVCG--DSRAGR[illegible]YG----TIACN----GCKGFF[illegible]SIWE-----QRDYVCRFG-----GKCLVVQE------YRNRCRAC[illegible]LRKCFT-V[illegible]MDAR-----
46|Q18646/1-82    ----KGERCVGCGS-ESAIRV[illegible]YG----ASSCH----GCKAFF[illegible]SVFE-----GRVYMCSAD-----NHCDITNE------SRNRCRAC[illegible]LRNCLE-G[illegible]MNPK-----
81|O17081/1-81    ----LTERCKVCED-RSNN-SRYG----APACL----GCTVFF[illegible]AVVN-----KMQYRCLRE-----QKCLISFT------YRCACRYC[illegible]FQKCLN-V[illegible]MRPN-----
28|O45328/1-81    ----PTENCKVCGD-RPNN-DRYG----ASACL----GCTVFF[illegible]AVVK-----NLKYKCMKE-----QNCLISCT------YRNACRYC[illegible]FQKCFA-V[illegible]MHSS-----
22|P90822/1-80    ----LSENCAVCGD-RVHS-VRLG----SPACL----GCIVFF[illegible]ATIN-----VSKYKCLKD-----GDCLINYE------FRSSCRAC[illegible]LQKCLQ-V[illegible]MKQS-----
13|O17013/1-81    ----LTERCAVCGD-QVKS-DRLG----CPACL----GCILFF[illegible]SVLK-----NASYKCGRK-----GACPVVSE------FRSTCRHC[illegible]EKKCFD-V[illegible]MKAS-----
46|O16346/1-81    ----SSESCAVCGD-SVNG-KRYG----APACL----GCIVFF[illegible]AVIN-----KSQYKCWKK-----GNCVITFA------SRCVCRCC[illegible]LRKCSH-V[illegible]MKPE-----
87|O45987/1-81    ----PQKRCLVCGG-TPNG-SRYG----ALACL----GCIVFF[illegible]AISN-----ANVKNCDNN-----SNCKIGFE------TKNCCRSC[illegible]LQKCLQ-V[illegible]MNPK-----
NH19_CAEEL/1-8    ----VIGQCLICG--EPSTGK[illegible]YG----IVACL----GCKTFF[illegible]AVVQ-----RQDTECKR--E---KPCDVTT------MARKACRAC[illegible]YRKCLE-S[illegible]MSKE-----
49|O45449/1-80    ----ENQPCMVCG--EISYSIRFG----AVSCR----ACAEFF[illegible]KIVSK---ARIPFRCN-------GACDLGKY------HRKTCQSC[illegible]FQKCLK-I[illegible]MLEK-----
28|O76828/1-81    ----TSQTCLICG--DSADSL[illegible]FG----ALSCR----ACAAFF[illegible]KVAGR---RNIFERCD-------RQCKVDTG------MRKLCASC[illegible]YDKCLK-V[illegible]MRES-----
1350602|E13506    ----EKPPCSICG--EVGNGV[illegible]FG----AEACR----ACAAFF[illegible]SVAL-----NKLYRCR--GD---GNCDILS------TIRCMCRAC[illegible]FAKCMA-V[illegible]MKRE-----
05|Q17905/1-81    ----GKSPCSVCG--EAGDGA[illegible]FG----AEACR----ACAAFF[illegible]SVAL-----NKAYVCR--AM---GTCVIQK------NVRCMCRAC[illegible]FTKCMA-V[illegible]MRKS-----
01|Q21701/1-81    ----EKPNCAVCN--EVGDGL[illegible]FG----AEACR----ACTAFF[illegible]SVAL-----SKKYECR--AG---RNCEVSS------NIRCMCRSC[illegible]YDKCLE-V[illegible]MNPE-----
4139088|G41390    ----VFLKCAICQ--ESAEGF[illegible]FG----AEACR----ACAAFF[illegible]TVSN-----RKTYSCQ--GN---NDCDVTI------NIRCMCRAC[illegible]YIKCLE-V[illegible]MNPA-----
3844612|G38446    ----NQSKCSICQ--ESGDGF[illegible]FG----AEACR----ACAAFF[illegible]SISE-----NKLYTCS--GN---NDCDVTV------NIRCMCRAC[illegible]LTKCLE-V[illegible]MNPV-----
33|O17933/1-88    ----PPSYCLICC--EVADGH[illegible]FG----AAACR(6)LACAAFF[illegible]TVQL-----NKVHDCP--KN---GQCFILS------NVRNMCRAC[illegible]YEKCLE-V[illegible]MQRS-----
34|O17934/1-97    ----PIPYCLICS--EVADGN[illegible]FGV---AAACR----ACAAFF[illegible]TVQH-----NKIHECG--RN---GQCFFLSCK(13)DARSMCKAC[illegible]YGKCLE-M[illegible]MQRS-----
00|Q21700/1-81    ----KKPNCAICG--DVGDGH[illegible]FG----LDACR----ACAAFF[illegible]TVAL-----GKKFECR--QQ---GACKINT------IIRSICKSC[illegible]LNKCLN-L[illegible]MKRS-----
25|O16425/1-83    ----ERKYCSVCH--QLGDGY[illegible]FG----AIACK----ACAAFF[illegible]TTSMNL---APKEVCR--KK---NECVIKM------SSRDSCKSC[illegible]YAKCLH-V[illegible]MNPE-----
55|Q18155/1-82    ----HRSTCKICG--LAAHGV[illegible]FG----VTACR----ACAAFF[illegible]FVVLN-----LEYECL-KDE---IKCNLNKI------RRSSCRHC[illegible]FQKCLR-M[illegible]NTAD-----
12|O62412/1-84    ----STLTCKICC--EPAHGV[illegible]FG----VASCR----ACAAFF[illegible]TTVLG-----KKSTCR-QKN---EKCTTNNE------DRFNCRYC[illegible]LQKCLS-N[illegible]MTPE-----
48|O18048/1-84    ----TIVLCKVCA--LSAHGS[illegible]FG----VLACR----ACAAFF[illegible]TVVLER---QKQYKCK-KGK---NECPVDTA------ERYQCRLC[illegible]RNKCVD-L[illegible]MTPE-----
1349031|E13490    ----STAICKVCG--LCAHGL[illegible]FG----VLACR----ACAAFF[illegible]TVVMER---QKKYKCR-GGE---DNCAVSST------DRYQCRLC[illegible]RNKCVE-L[illegible]NTPE-----
56|O45756/1-83    ----IAAKCRVCA--MPSRGN[illegible]FG----VLSCR----ACAAFF[illegible]SFVHK-----KEYKCRQRNV---GKCDVSGN------DRYQCRAC[illegible]LKKCKE-L[illegible]MTPE-----
```

Fig. 19 continu d:

```
38|Q23038/1-82    ----PKQKCRVCN--LEAHGMFG----VQTCR----PCAAFFIVVLD-----LKYTCV-SNT---QKCNVDGR------GRNVCRDCYKKCIA-VMTTD-----
89|Q20389/1-82    ----PDKSCLVCK--KASNGMFG----ALTCR----ACAAFFAIVLK-----LQYKCK-QGN---SKCNIDGR------GRYVCRQCLTKCKA-IMNEE-----
1354403|E13544    ----FTVNCEICQ--KTSHGLFG----LETCR----ACAAFFTVVLN-----RKYKCI-QKT---GKCEIGAGG-----EEKLCKFCHKKCED-LMTTE-----
10|O45910/1-81    ----TPDKCEICQ--KTGHGHFG----LETCR----ACAAFFTVVLN-----RKYKCA-QKS---GKCDVGA-------EKATCKFCYKKCED-LMTTD-----
96|Q19496/1-88    ----ISMSCLVCE--TDAHGQFG----IRCCR----ACAAFFTITMN-----LRYKCR--FD---RKCEVSFN------KRYSCRCCYEKCVR-VMRKD-----
62|O16962/1-81    ----SPDTCQVCG--QKSHGKFG----AVTCR----ACAAFF----CGSANNFKP-CRR-NM---N-CEFLKN------GWFNCKPCLKKCQD-VMTIE-----
67|O16667/1-83    ----NQTFCQVCG--QESHGAFG----AITCR----ACAAFF----VAAGANFEVKCKD-GR---GRCKILTN------GRSCCKKCLKKCKD-IMDIQ-----
05|O45905/1-84    ----SELICAVCS--QPARGRFG----AVACR----ACAAFF----ADAAKTTVKPCKK-GG---N-CQNLLNNN----GWFDCKFCLQKCYE-IMNSA-----
63|O16963/1-82    ----NLYTCQVCA--LPAHGNFG----AISCR----ACAAFFA---CTGTKTT-YKCKK-QN---N-CDIWEN------GRYKCKKCLDRCNE-VMDPG-----
64|O16964/1-84    ----SQILCQICA--LPAHGNFG----AVSCR----ACSAFFS---ILEKNAENYKCLR-GN---NTCEVKLA------GKFYCKKCHKRCYE-TMDPK-----
55|Q22555/1-83    ----ISEKCLVCD--QPSHGNFG----VDSCR----ACAAFF----VFVTHKQQYKCIN-EL---EQCVPDAR------GRWKCRKCTDRCFE-LMRPD-----
55|Q22555_1/1-    ----ISEKCLVCF--QPSHGNFG----VDSCR----ACAAFF----VFVTHKQQEPCRE-GD---NKCTPDEW------GRWSCKRCSDKCFA-LMKPD-----
56|Q22556/1-81    ----PIIKCKVCY--QPAHGNFG----VEICR----ACAAFF----IFVSHKQHEKCLD-GR---KKCTPDAA------GRWTCKRCENRCYA-LMKPD-----
52|O16752/1-83    ----KSRKCQVCY--LRAHGNYG----VVSCR----ACAAFF----VFITHKQHEKCQG-GQ---MKCEPNSV------GRWDCKSCTKKCFD-IMRPD-----
03|O61203/1-83    ----PRKDCRICF--EKGHGYFG----VFSCR----ACAAFF----CHFLKENRERCRL-SK---GKCGPNKN------GKWFCKTCYERCFR-LMTPS-----
37|O44637/1-83    ----IHEKCRVCD--QPGHGFFG----AFTCR----ACSAFF----AHFSLVTEEKCRL-SN---GRCVPTKG------GRWFCKKCLAKCSE-AMNSR-----
53|O16753/1-81    ----FDLKCEICC--SKGHGNFG----VTACR----ACAAFF----MIVGTGYRQKCR--SY---KTCEP-KD------GRWQCKKCLEKCYS-LMTPD-----
57|O16357/1-87    ----RKSLCKICG--NPAFGNFG----VLSCR----ACAAFF----TNVDG--ETFCPI-GN-----CKIDKH------GKFYCKKCYRICIE-AMDAN-----
54|O16754/1-83    ----DMSCCKICG--IQAHGKFG----ALTCR----ACAAFF----CHFNTGPEIVCQF-MN---SKCKADNK------GRWACKKCFDRCIA-NMTTR-----
58|O16358/1-79    ----LSGPCEICG--QKTSGRFG----VMSCR----SCAAFFAI----TWKTERADCKK------ENCQIFERG------KFSCKVCLKRCIE-VMDAS-----
61|O16361/1-81    ----LSGPCEICG--QKTSGRFG----VMSCR----SCAAFFAI----TCSRIFGRCPN------GNCKLLENG------KFKCKKCMKRCIE-VMDET-----
F44C8/4-0         ----LSGPCEICG--QKTSGRFG----VLSCR----SCAAFFSI----TWSRKKVQCVK------GTCKIFEDG------KFNCKQCLKKCVE-VMDSK-----
4139090|G41390    ----LSGSCEVCG--DKTSGRFG----VMSCR----ACAAFFAI----TWNLEKEICPN------GTCHTSVNG------KFNCKQCLKKCLD-VMDTR-----
60|O16360_1/1-    ----LSGPCKICG--QNTSGKFG----VLSCR----ACAAFFAI----LWSPKKAHCPR------GNCASFENK------NLNCKQCLKKCLD-VMDIS-----
F44C8/11-0        ----SFGLCAVCG--QVTTGKNFG----VISCR----SCAAFFAI----TWSRRFPECVK------ASCAIFENG------KFKCKKCLKRCQE-VMDVK-----
F48G7/3-0         ----LSGPCVICD--QPSHGRFG----VLSCR----ACAAFFAIR--RSAKSQDMVCER------ENCSLEDG-------IQYCKICLKKCFN-AMDPN-----
F48G7/11-0        ----LSGPCVICG--QPSHGRFG----VLSCR----ACAAFFAVM--RSAKSQDMVCER------ESCSLKDR-------IQCCKICLKMCVD-AMDPN-----
54|O16354/1-81    ----LSGPCEICE--QPAHGNFG----VLSCR----ACAAFFAIL--QNAKYQDAVCRK------GNCIGND--------LYRCKICLKKCCE-VMNSS-----
55|O16355/1-80    ----ISGPCKICD--LPAHGNFG----VLSCR----ACAAFFAI---RSPRFQDKACEY------GDCRIFDAG------AYRCKICLKKCYK-VMDAS-----
12|O01612/1-89    ----ILFPCRICG--KKAHGTFG----VFSCR----ACAAFFLVTSLILGMAISFRSCME-IF---VKCLAKGGSG----NECLCKPCLKKCFE-TMDPS-----
F31F4/12-0        ----SHKTCQICKV-EKCHGNFG----VDSCR----ACAAFF--------RTVHSRWSRE-------KCQGGTCDR----KKLLCKPCLQQCFE-GMVTD-----
61|O16961/1-82    ----NGPECEVCG--LESQGFFG----VLSCR----ACSAFFTAT---VPKWHQKNCQK-P----SKCKNGEG-------QYHCKPCLKRCIA-VMNTE-----
66|O16966/1-85    ----MQTTCEICE--VPAHGIFG----AITCR----GCAAFF----AVNTKTNRKSCKY-S----SNCTNFTG------KFPQCKSCMRKCIK-MMMPE-----
65|O16965/1-79    ----MSKKCGICD--APAHGRFG----ALTCL----SCAAFF----AIFAKKKRMSCKS-G----NKCRDFNG------NFPQCKSCMRKCIR-VMKLE-----
NH22_CAEEL/1-8    ----DSRSCHVCSS-PTANTLFG----GRS--(1)KACAAFFSVSM-----SMTYECIGTGDD-TNPCRTHY------ELRMICRHCFIKCID-AMRRE-----
94|Q23294/1-82    ----MGPLCAVCES-PTAFTLFG----GRC--(1)KACAAFFTIAL-----DLKYECA--AD---DPCEIHF------SMRLVCRECLRKCYS-AMRSE-----
41|O18141/1-84    ----VRGKCMVCDS-PNATNYFG----AQS--(1)KACAAFFSIAM-----DQCYECLGDGV---HSCKIDH------TLRLNCRHCLKKCQK-AMMRD-----
77|O44577/1-82    ----SRGNCKVCDS-PNATNFFG----AHS--(1)KACAAFFTVKT-----GQIYSCPGE-----KTCEINH------VFRQNCRQCHAKCIK-AMQEE-----
4139082|G41390    ----NPNSCEVCSSSSNSSCNFG----ART--(1)KACAAFFTVSM-----KLDYQCIDQP----DACRVHC------DSRVICRFCLKKCED-IMKPL-----
ODR7_CAEEL/1-8    ----ALHDCQVCLS-THANGLFG----ART--(1)AACAAFFTISD-----DKRYVCKRNQRC-NNASRDGT------GYRKICRSCMKRCIE-IMLPE-----
68|O76668/1-82    ----SSTRCLICSA--QATGFFE----AQS--(1)SACAAFFTVAL-----TKSFKCIT-GR---DDCNVHY------SMHQICRSCHKKCLS-NMKPG-----
3886065|G38860    ----SPHPCFVCNS--PANECFG----SRSQI(1)RACSAFFYVLSP----KSADKCRK--D---KSCRIHY------KDPKICRFCMERCVL-AMKTT-----
91|Q18391/1-75    ----YAMLCKICDG--PSTEYFG----CPS--(1)RACAAFFYVNSP----KQLTECN----------CETD-------RVP--CRYCMKLCLK-AMMIS-----
87|O18087/1-82    ----LSIPCEVCKN--QSNGYFE----VLS--(1)GACASFFSVVS-----KIKYQCKDGK----KRCQIRY------LDRHFCRYCESKCVK-AMKAE-----
NH20_CAEEL/1-8    ----PTSKCLVCEHPDGGSA-FG----STS--(1)LACAAFFTVSL-----NIQFQCKKD-----KNCVIFHEL------RMICRACEDKCVK-AMRRE-----
1348628|E13486    ----VARTCLVCTITENVRF-FG----STT--(1)LACASFFTVSL-----KIQYVCKQS-----NNCIVSHAV------RSGCRSCFQNCLK-SMKTN-----
84|O16884/1-83    ----STPTCLVCGLSCRTHF-FG----GAS--(1)SACASFFTVSL-----NISYLCKRD-----SDCISSPEL------RSICRACYQRCEDTANSRD-----
12|O62112/1-83    ----PNSKCLICGSICTAKF-FG----GLS--(1)NACASFFTVSL-----NIRYLCKNS-----NNCITSPGQ------RTFCKACYESCVDKAMKRD-----
```

Fig. 19 continued:

```
16|O02316/1-80    ----ADPICPVCEFPSNVEL-FG----GLV--(1)GACAAFFTVSL-----NIRYLCEKN-----NQCKGMRKN---------CRACFDYCVKIAMKRN-----
06|O17706/1-80    ----TDPICSVCNFSSLIAP-FG----GLV--(1)SACASFFTVAL-----NIHYLCKKD-----NQCKGMRKN---------CRACFESCVKIAMKRS-----
07|O45507/1-85    ----EFTHCPVCSLPCSMRYSFG----GIC--(1)PACASFFTVSL-----NIRYLCKKQ-----NNCSGISKKY-----HIVCRACYEKCIKKAMKRS-----
68|O44668/1-82    ----FSKPCQVCSTSERTSY-FG----TTT--(1)MACASFFTVAF-----GIRFLCKSN-----NNCSISPKA------RFFCRSCFDKCVN-VMNKS-----
1347658|E13476    ----QYDPCQVCQDTASTRH-FG----ITS--(1)TACASFFTT-------MNQFTCNLD-----NNCSVSY------KKKWVCRACYNNCIR-AMSKD-----
04|O16604/1-82    ----SYVPCLVCNDVKNTSM-FG----VNS--(1)AACAAFFTT-------ENQYACSRN-----NNCATSY------ERKILCRACYTNCLE-AMSRG-----
1354720|E13547    ----KNDPCRICNDSLTSGF-FG----TSC--(1)IACAAFFTV-------KNQFTCNAN-----NDCVISF------EKKFTCRACYNNCLK-AMRKD-----
1354390|E13543    ----EKAACQVCQSTERTRL-YFG----ITS--(1)AACAAFFSD-------GIQYMCTAN-----NSCTIAYGR(18)DIKFFCRACYESCVR-AMRRD-----
1354721|E13547    ----EIASCQVCHSTERTRL-YFG----ITS--(1)AACAAFFSD-------GIRYMCTAT-----NSCTISH------DMKFFCRACYTSCIR-AMVMS-----
05|O45505/1-80    ----NRGPCPVCSSMDRTRR-FG----IIA--(1)TACAAFFSF-------GQQYLCNAE-----KKCIIMH------DRKFFCRACYSSCVK-ANRRD-----
16|O62116/1-76    ----NRGPCQVCHNTETTRR-FG----IIS--(1)TACAAFFSI-------GMQYLCTAE-----NRCTISF------DRKFFCRACYASCLK-AMKLE-----
35|O02235/1-80    ----PEGPCRVCHSVKGTRR-FG----ITA--(1)MSCSSFFSL-------NCKFYCPAN-----NSCTILD------DQKQFCRSCYNKCVQ-SMRRD-----
T09A12/4-0        ----AVPACAICG--TDSTGIFG----VDACA----ACSAFFTVVI-----NKDYSCNK--G---GKCTVVKDG----SAGQKCRACFRKCIS-SMDKN-----
17|Q19217/1-81    ----DLGNCKICL--QRADGIFA----VSSCR----ACAAFFTVII-----KLNYTCKE--K---GNCTVEK------SLRNLCRSCYTRCIN-EMKIE-----
31|O01931/1-85    ----CPSKCLVCR--NPAIGYYD----VPSCN----GCKTFFTIIT-----GRKFTCAKQ-----KKCMDGTEPVD--MSKRLCRACYAKCVE-VMNPM-----
29|O01929/1-85    ----CPSVCQICR--NPAIGYYE----VPSCN----GCKTSFTLIT-----GRKEKCFKV-----SNCLDGNDVID--TSKRVCRACFEKCVQ-AMNPM-----
30|O01930/1-85    ----CPSNCKVCR--HSATGYYD----VPSCN----GCKTFFSILL-----GRKYTCLKM-----RKCLSGTEPVD--LSRRMCRACFEKCVE-AMNPS-----
06|Q21806/1-84    ----FPEKCAVCK--NAAIGYYN----VPSCN----GCKTFFTILN-----GKRFICMNH-----KNCLDEIE-SD--ESQRLCKGCEARCIE-VMDST-----
03|Q21803/1-83    ----LPEKCEVCG--NPAVGYYD----VATCN----GCKAFFTVIT-----GRRVICRRR-----KNCLVE-MEPK---NRRICPGCFSKCIE-VMNPR-----
05|Q21805/1-83    ----LPSKCEICG--NPAIGYYD----IASCN----GCKAFFTVIT-----GRQVACKKW-----GTCLEE-EIPI---NRRICPGCESKCVK-VMNPR-----
04|Q21804/1-85    ----LPKQCRICR--NPAVGYYD----VASCN----GCKAFFTVIT-----GRLPNCKFG-----DKCLEDHKIPK--PGTRLCGSCYSKCIQ-MMNPM-----
48|O01448/1-84    ----TPITCSVCG--RSAKGYYD----VISCK----GCKTFFMYLL-----KSKDICGLN-----NSCFDLKNKN---EPYLRCRACYKKCID-VMNRN-----
46|O01446/1-85    ----IPTTCSICR--RNAKGYYD----VISCK----GCKTFFMSLI-----NFQGKCDLN-----NQCFDLTKAN---EPYLRCRACYRKCID-VMNQK-----
40|O61940/1-84    ----KPSKCMVCA--RPAHGYCD----VTTCK----GCKSFFMYLL-----RSETKCSGR-----NDCFDLKKRI---EPLLRCRSCYRKCIL-VMNPE-----
29|Q17929/1-84    ----SPFNCKICN--RPAHGHCD----VATCK----GCKTFFMYLL-----KSELRCYIN-----SDCFDLDKRT---EPLLRCRACFKKCIL-VMNPN-----
49|O01449/1-85    ----FSHLCSICS--RPAQGYYD----VISCK----GCKTFFMWLS-----KIKEMCPLN-----NKCFDFNRRIN--MSLSKCRACFQRCIN-VMNPA-----
83|P91983/1-83    ----RPSECSICG--KTANGYYD----VPSCN----GCKTFFLCVS-----EKSFICKAG-----GDCFNLTKRK----VPLKCRACYEKCIS-EMNPN-----
98|O17898/1-83    ----RPTECLICG--NSANGHYD----VASCN----GCKTFFMGVS-----DRSFECKAK-----GDCFDLTKRK----VPLKCRACHQKCIS-VMNPL-----
1345476|E13454    ----RPTECLVCG--RSAHGYYN----VASCN----GCKTFFMCLS-----GRSFSCKLD-----RDCFDLTKRK----TPAKCRACLQRCIS-VMDPG-----
83|P91983_1/1-    ----RPKDCFVCG--GSASGYYD----VPACN----GCKTFFYVIS-----KRLATCKLN-----DDCFDLSTRD----VPVKCRACINKCFA-VMNPH-----
43|O44543/1-84    ----VPSACQVCY--GKATGHYD----VASCN----GCKSFFVTIE-----KLKYRCSAG-----NACYTNFGPG---DPVPKCRACYKRCIE-AMNAL-----
16|Q20916/1-79    ----VPKLCEICN--RQAFGKYG----VLSCY----ACKMFFMSVQ-----KLVYNCKWS-----NNCFETSP------YPLKCQSCFQRCID-AMYLE-----
C54F6/8-0         ----FEKVCTVCR--RPAFTRYYG----AVSCE----ACKMFFMIIE-----KLSYDCKKI-----NNCYETSKG-----TRFKCRSCFKRCIQ-MMKLP-----
27|O17927/1-97    ----MNKNCLICH--RKAAGQYG(16)VLSCF----ACKMFFHMVVE-----NLHYCCQKF-----NKCYEKFI------ILPKCKACYQKCIE-MMQAF-----
62|O01562/1-82    ----STLKCLICG--RNASCHNYG----VLTCN----ACKMFFVVIE-----QITYQCRNW-----KKCLENADQ-----FKPRCKACFEKCVR-LMKLG-----
63|O01563/1-80    ----NAPKCTICG--RDASCHNYG----VLSCN----ACKMFFVVIE-----QLTYHCKKW-----NMCLENSGQ-----PILQCKACFEKCIQ-LMTLG-----
79|Q20379/1-78    ----QDKKCTICS--RPASTRNHG----VLSCS----ACKMFFATVQN-----KLKEECKKE-----NQCLGNFSH-------RICQSCFEKCIK-LMRCS-----
64|O45664/1-73    ----VREKCVICF--GSALGMNYG----ALTCS----ACKMFFSAEK-----KLSYNCKKD-----NQCSIAQRK---------CRACFQKCLN-FMTFY-----
T24A6/11-0        ----PSKQCSICQ--RPAFGQNYG----SFSCD----ACKMFFCIML-----DLQFECIYQ-----KRCFDK---FTEHRRVPRCKACFQKCID-TMFIR-----
F47C10/1-0        ----CSIQCSVCC--QLAFGLNYG----VPSCN----ACKMFFMNII-----KPSYTCLAD-----GSCYTN---FGFTHGRPKCRSCFRKCID-VMKYR-----
F47C10/3-0        ----VSEKCAVCG--KPAFCYNYG----VLSCN----ACKMFFVEVD-----HITYNCKYW-----NTCYDGQEYVDANNVSTRCRACYRRCID-LMRYI-----
54|O61854/1-85    ----SSEKCAICG--NLAFCYNYG----VLSCN----ACKMFFVEVD-----RITYHCKYW-----DRCYDGQEFVNVKKFSKRCRACYQRCLQ-VMKYI-----
89|Q23489/1-80    ----SERRCATCS--KLGNSYNYG----VLSCN----ACKMFFATLG-----KSEKFCINN-----DNC------(3)-NLILTCRQCYNKCIK-MMRIV-----
09|O16609/1-82    ----KPFACAICR--NKSSGCYG----VLSCD----ACKMFFKTYLM-----EKYYTCRRQ-----ESCYNMDGG------NRKCKACFKKCIE-VMNPE-----
97|Q22297/1-84    ----ITRSCKVCN--EKTTGFNYG----VQSCN----ACKMFFALLT-----PKVEKCGRD-----GNCPTIVLDSG----ESQCRFCFHKCII-VMIFL-----
31|Q20831/1-78    ----KHSFCDICG--KKSFGFNYG----AHSCN----ACKMFFALST-----RKIYQCSKN-----GNCSTIAVET-----EKFCRFCFHKCIL-AMILV-----
T24A6/8-0         ----KPSKCMVCVR--PAHGYCD----VATCK----GCKTFFMCILR----G-EIKCSTS-----GDCYDLEKRNS---PLLRCRPCFKRCIL-VMNPK-----
60|O45460/1-84    ----FSVKCAICY--KAGHGQFG----VETCR----ACAAFFTVVLN-----RKMKCTRKS----GKCKIGSDE----TKDVMCKFCEKKCID-LMTTE-----
```

Fig. 19 continu d:

```
FXR-B            (17)QEEFCVVCGDKASPSPYFYN-ALTCEGCK(10)FLLGFFQCSIX-----QNAVESCRNG-----SHCEMDMY------MRRKCQECRLKKYHA-VGMLAE(279)
57|O17657/1-80   ----KFGVCAVCGF-SCQVQYHFG----GVVCG----ACSAFFRVTVSLN-----IRYLCDGD-----NQCKSMLK---------KCRACRYESCVKTAGMKRS-----
4139078|G41390   ----AGRICTVCSD--RANGYNFG----VLTCE----SCKAFFRRNASK-----HKEIKCPFS-----DSCQITSA------SRKFCQACRLNKCFA-VGMNSE-----
65|O45365/1-80   ----TIQPCLVCG--QSSNSILFG----APSCR----ACGEFFRRKVISN---FKIKNNCLG-------ECSFAKK------SMKPCQSCRFQKCLE-AGMLEK-----
38|O16538/1-96   ----KNGPCRICND-PETTSPYEG----AIVCT----ACASFFGRTTVD-------KERCIAN-----NKCVISCGE(14)AKKITCRSCRYDSCIR-AGMNKF-----
68|O16968/1-78   ----SCPPCIICGQ--NGQGHYFG----VISCR----ACAAFFRRAADS---KWSRMKCLSK------YCDGKTY---------HCKPCRLKKCRD-VGMDVS-----
86|O18086/1-79   ----QPIPCQICTY--QSHGVNEN----VMTCR----ACAAFFRQVIKI---------VHCKTRK----NDCRIDST------ERHFCRLCRFQKCLQ-MGMKAE-----
91|O16391/1-87   ----RPNICAICH--QKAFGYNYE----VVSCN----ACKMFFRRAHAE-----KIDDFCKKG-----GKCFDGDDLLT---SRPKCRSCRYKKCVN-LGMRYH-----
55|Q17755/1-10   ----DGPCCLVCGR-VANTGHYYG----VTACL----GCKTFFRRVVLQ-----KNSPKCKYK-----NHCRLEKSM(17)NAKRLCRSCRYQKCLE-VGMTED-----
34|O62034/1-78   ----TSTSCHVCG--APEVEPYFG----GVSCR----ACAAFFRRYFHSK----KSVISCT----------CQIRYQN------SHPCRECRLQKCYE-AGMNPE-----
03|O46003/1-81   ----LFEACAVCGG--PPSSRRFG----ALACF----RCNVFFRRNMLLN-----KRNIRCSKN-----PECEIGIK------TRYACQGCRFKKCEK-SGMDPK-----
3800990|G38009   ----RFNPCDVCLS--PGNGKYFG----VDACR----GCTAFFRRTIVN-----KRNYSCSED-----DTCEIEKS------EGLLCKKCRFEKCLK-VGMKKE-----
90|Q18190/1-82   ----QFSICRVCDD--SNGQPYYG----TICCP----SCKGFFRRVYMS-----DKKYECYRG-----NMCVIQKG------TRNICRACRYKKCLQ-VGMNVK-----
CAA05409         ----ARRLCLVCED--YASCSNTC----VWSCE----AYKVFFRRSQS------FTDPACFTN------DCNISKN------RSKSCPACLLRCLQP--SINEI-----
ZK6/5-0          ----SNQTCAVCER--FTTEFNYG----VPSCN----ACKIFFRRLITR----TAPVKQCYIG-----EHCFTKSP------ITKKCTFCRFQKCIQ-VGMTLP-----
55|Q19155/1-83   ----SNEMCRVCG--KNSAGKYYS----VPACH----GCKSFFRRATIHK----TAYPECKYD-----KMCFSKISR---ITRPRKCRYCRLQKCYE-VGMVAI-----
64|O01564/1-84   ----GQKACVVCKR--PANCYNFG----VMSCD----ACKMFFRRTMLL-----NVQYVCRRK-----NQCFDDSFS---CLKTPYCRACRLVQCVK-VGMKLN-----
11|Q94411/1-85   ----KVENCLICDR--ETSLFNYG----AISCN----ACKLFFRRAKVS----KTEISPCSNR-----NLCHKTKMS----TWKSACKSCRLLKCVK-AGMTLN-----
99|Q18299/1-84   ----SNIHCRVCER-RYDGSQYFG----IDICR----ACAAFFRRSVAV-----KKTEVCRRG----TNKCELNTVS-----RKTTCQKCRWMRCEL-VGLNVD-----
Consensus/80%    ........C.lC.....s...Hat....s.sC.....tC.sFFRRsh.........b.C.........pC...........p..Cp.CRbp+Cbp.hGMp.p.....
```

Fig. 20:

```
05|Q17905/1-18   ---NVHLKEIGLVADWIIKSYPDHE--QLHQEQKKLLYRNFFLPFMI-LECG----YMCCLNNRTDILFLPS-----------------GDYIDCNRPETFYGHR-----IKSHLISPNE
AC68990|AAC689   ---RLLLREFYLVSDWVSNCFPLEK--ELPNSQKDILLRNFYLTFYN-LEAG----FFACQHNKIDVWFLPN-----------------GNFIDCQNLEAFYYDP-----HNQQPMTSAD
01|Q21701/1-18   ---EVTTKEVSLVADWIEWCFDDFG--LLPVDQKTILFQNFFVFFCM-LERA----FMTVKSGRDGIVVMAS-----------------KDYIDYDNLEGFFKEC-----NSGTGGTPAE
33|O17933/1-18   ---VVYQNEMTVIYQFLCEAFPEYS--ELLPDTKRSLKKNFWLPFTL-LESS----YYGYLTKQENVMLIPS-----------------GDYVDLAHLESYFN-------NNHHTFSQKD
92|Q18392/1-17   ------NFNTNLLWKLGESIFSDVS--RLSKEDKNSMISNFYIKWQL-LMEP----AIDYSAQYEQFKNYIG-----------------SNAYYQKCAYFYGSSM-----QKGNKISDVE
95|Q18395/1-18   ---SFVKRDTHLLWDLGEKIFTDVK--LLSEADKHSILCNFFPRWMM-LDS-----AVAICPDYEESKAYIK-----------------SKDYTDMLLHFYGSSM-----PKEKRLKDHE
96|O62496/1-17   ---SFVLRDLELITWRMVSKAFPTVE--VLEDRDRQALLRNFVPKLWQ-LEP-----ILEYRECAEKFDNLE-----------------EEDYENLLVGFYGGTF-----PEGKEMSKSE
73|O17573/1-17   ---TSSAGDMDSTWKMVINLFSSLG--ELEIKDKTALLRNFMPKFIQ-LDS-----VPYFAANIDVFKNIG------------------RDEYESSIIDFYDGVL-----PETNTISKKD
31|O62031/1-17   ------IEDCMMIWRIVEDLFPSLTW-NLRKLDKEALLRNFLPKWSV-LTA-----AIDMEVKVQRYSKVNN-----------------TEEFKRMIVDFFTYSM-----PEKFKMPPED
49|O61449/1-16   ---CQAAP-PLHELVQWARRIPHFE--ELPIEDRTALLKAGWNELLI-AA-------FSHRSVAVRD-----(1)-------------IVLATGLVVQRHSAHGAG--------VGDIFDR
48|O61448/1-15   ----QAADRQLHQLVQWAKHIPHFE--ELPLEDRMVLLKAGWNELLI-AA-------FSHRSVDVRD-----(1)-------------IVLATGLVVQRHSAHGAG--------VGAIFDR
41|O76241/1-16   ----QAADRHLVQLVEWAKHIPHFT--DLPIEDQVVLLKAGWNELLI-AS-------FSHRSMGVED-----(1)-------------IVLATGLVIHRSSAHQAG--------VGAIFDR
15|Q90415/1-16   ----QAADKQLFTLVEWAKRIPHFS--DLPLDDQVILLRAGWNELLI-AS-------FSHRSIAVKD-----(1)-------------IILATGLHVHRNSAHTAG--------VGAIFDR
62|O96562/1-16   ----LAADKQLASLAEWAKRLPHFR--DLSIADQVVLLQWSWPELLI-GG-------FCHRSCAVKD-----(1)-------------ILLSTGLHLTRDNLKKAG--------VGAIIDK
53|USP_DROME/1   ----QVVNKQLFQMVEYARMMPHFA--QVPLDDQVILLKAAWIELLI-AN-------VAWCSIVSLDDGGAG(7)HDGSFER(7)QQLFLNQSFSYHRNSAIKAG--------VSAIFDR
02|O76202/1-17   ----QIGNKQIAALVVWARDIPHFG--QLELDDQVVLIKASWNELLL-FA-------IAWRSMEYLEDEREN(3)------T(6)QLMCLMPGMTLHRNSAQQAG--------VGAIFDR
35|HN4A_HUMAN/   ----ESMKEQLLVLVEWAKYIPAFC--ELPLDDQVALLRAHAGEHLL-LG-------ATKRSMVFKD-------------------VLLLGNDYIVPRHCP--ELA------EMSRVSIR
26|HN4B_XENLA/   ----ESMKQQLLLLVEWAKYIPAFC--ELPLDDQVALLRAHAGAHLL-LG-------VAKRSLPYKD-------------------FLLLGNDFIMPMHCP--EL-------EIARVPCR
66|HNF4_DROME/   ----ESMKQQLLTLVEWAKQIPAFN--ELQLDDQVALLRAHAGEHLL-LG-------LSRRSMHLKD-------------------VLLLSNNCVITRHCPDPLVSPN---LDISRIGAR
75|O46175/1-16   ----DSIKQQLLILVEWAKYIPAFT--VLHLDDQVALLRAHAGEHLL-LG-------CARRSLHLKD-------------------ILLLGNNCIITKHNIDGRMD-----IDISMIGMR
48|FTF_MOUSE/1   ----KMADQTLFSIVEWARSSIFFR--ELKVDDQMKLLQNCWSELLI-LDHIY(5)-GKEGTIFLVT-------------------------GEHVDYSTIISHTE------VAFNNLLS
45|Q91545/1-16   ----KMADQTLFSIVEWARSSIFFR--ELKVDDQMKLLQNCWSELLI-LDHIF(5)-GKEGSILLVT-------------------------GQQVDYSVIVTQAG------ATLNNLMS
58|O93258/1-16   ----RMADQTLFSIVEWARSCIFFK--ELRVGDQMKLLHNCWSELLV-LDHIF(5)-GKEDSILLVT-------------------------GQEVQLSSILSQGE------ATLCSLVQ
54|Q9YI54/1-16   ----KMADQTLFSIVEWARSCIFFK--ELEVGDQMKLLQNCWSELLV-FDHIY(5)-GKEHSMLLVT-------------------------GQEVEMATIAAQAG------SNLNNLVL
44|FTF1_DROME/   ----KVLDQNLFSQVDWARNTVFFK--DLKVDDQMKLLQHSWSDMLV-LDHLH(5)-GLPDETQLNN-------------------------GQVFNLMSLGLLGV------PQPGDYFN
41|O96041/1-15   ----NIADHRLYKIVKWCKSLPLFK--NISIDDQICLLINSWCELLV-LSCCYRGVS-TPGEVRVGG-------------------------GRGITLHQSAKLG---------LTPCIE
36|O96836/1-21   ----NIGDSIVHKLVSWTKKLPFYL--ELPVEIHTKLLTDKWHEILI-LTTAA(29)GTPLSTPIPSPAQPL(5)----EFV(7)STLQTCLTTLMGQPIAMEQLK------LDVGHMVD
28|O61228/1-18   ----HIGDSIVYKLVQWTKRLPFYL--ELPVEVHTRLLTHKWHELLV-LTTSA(6)AGDQLTTVIKT-----------DFN(7)CTLQSCLTSMMGREITIEQLR------QDVGLMIE
33|P70033/1-16   ----RLADELLFRQITWVKKLPFFC--DLSIKDYTCLLSTTWQELIL-LS-------SITTYSKQIFG-----------------DLTDVTSKYSPSEDELHRFS------EDGMEVME
18|COT2_RAT/1-   ----ELAARMLFSAVEWARNIPFFP--DLQITDQVALLRLTWSELFV-LN-------AAQCSMPLHV------------------APLLAAAG--LHASP-------MSADRVVAFMD
36|EAR2_MOUSE/   ----ELAARLLFSTVEWARHAPFFP--ELPAADQVALLRLSWSELFV-LN-------AAQAALPLHT------------------APLLAAAG--LHAAP-------MAAERAVAFMD
27|Q60927/1-16   ----ESASRLLFLSMHWALSIPSFQ--ALGQENSISLVKAYWNELFT-LG-------LAQCWQVMNV------------------ATLLATFVNCLHSSLQQ---DKMSPERRKSLME
16|TR4_HUMAN/1   ----ESASRLLFLSMHWARSIPAFQ--ALGQDCNTSLVRACWNELFT-LG-------LAQCAQVMSL------------------STILAAIVNHLQNSIQE---DKLSGDRIKQVME
22|Q26622/1-16   ----ESASRLLFLSMHWARSLPAFQ--VLSADTHTSMVQKCWSELFT-LG-------LAQCAQAMAL------------------STILTAIVNHLQTSLQQ---DKLSADRVKAVME
AD28301|AAD283   ----ETSARLLFMAVKWAKNLPVFS--SLPFRDQVILLEEAWSELFL-LG-------AIQWSLPLDS------------------CPLLAPPE--ASAAG-------GAQGRLTLASM
52|TLX_XENLA/1   ----ESAARLLFMSIKWAKSVPAFS--TLSLQDQLMLLEDAWRELFV-LG-------IAQWAIPVDA------------------STLLAVSG--MNNEN-------TESPKLNKIIS
20|O46520/1-17   ----EAASAGLLKTLRFVKYLPCFQ--VLPLDDQLVLVRSCWAPLLM-LE-------LAQDRLNFET------------------VETLEPSLLQMILTTRRQET(12)QPHLVLPSE
43|DAX1_HUMAN/   ----EAASAGLLKTLRFVKYLPCFQ--VLPLDQQLVLVRNCWASLLM-LE-------LAQDRLQFET------------------VEVSEPSMLQKILTTRRRET(11)QHHLAPPAE
AB45139|CAB451   ----SMADKELVHMIAWAKKVPGFQ--ELSLHDQVQLLESSWLEVLM-IG-------LIWRSIHCPG-------------------KLIFAQDLILDRSEGDC---------VEGMAE
72|ESR1_HUMAN/   ----NLADRELVHMINWAKRVPGFV--DLTLHDQVHLLECAWLEILM-IG-------LVWRSMEHPG-------------------KLLFAPNLLLDRNQGKC---------VEGMVE
12|O13012/1-17   ----NLADKELVLMISWAKKIPGFV--ELDLSDQVHLLECCWLEVLM-LG-------LMWRSVDHPG-------------------KLIFSPDLKLNRDEGSC---------VEGILE
86|ESR2_RAT/1-   ----KLADKELVHMIGWAKKIPGFV--ELSLLDQVRLLESCWMEVLM-VG-------LMWRSIDHPG-------------------KLIFAPDLVLDRDEGKC---------VEGILE
AA74855|BAA748   ----DLADRELVVIIGWAKHIPGFS--TLSLADQMSLLQSAWMEILI-LG-------VVYRSLSFED-------------------ELVYADDYIMDEDQSK----------LAGLLD
75|ERR2_HUMAN/   ----DLADRELVFLISWAKHIPGFS--NLTLGDQMSLLQSAWMEILI-LG-------IVYRSLPYDD-------------------KLAYAEDYIMDEEHSR----------LVGLLE
97|O93497/1-16   ----ELGERQLVKVVKWAKGLPGFR--NLHVDDQMTVIQHSWMGVMV-FG-------LGWRSYKNVNG------------------RMLYFAPDLVFNEHRMH----------ISTMYE
V9|Q9YGV9/1-16   ----ELGERQLVKVVKWAKGMPGFR--SLYVDDQMTVIQHSWMAVMV-FA-------LGWRSFKNVKS------------------RMLYFAPDLVFNEHRMQ----------VSTMYE
99|ANDR_RABIT/   ----ELGERQLVHVVKWAKALPGFR--NLHVDDQMAVIQYSWMGLMV-FA-------MGWRSFTNVNS------------------RMLYFAPDLVFNEYRMH----------KSRMYS
AD25074|AAD250   ----ELGERQLVTVVHWAKAIPGFR--DLYVDDQMSLIQLSWMGVMV-FA-------LGWRSYTLTNS------------------SMLYFAPDLVFNDQRMQ----------VSSMYE
```

Fig. 20 continu d:

```
15|GCR_CAVPO/1    ----MLGGRQVIAAVKWAKAIPGFK--NLHLDDQMTLLQYSWMFLMA-FA-------LGWRSYKQSNG------------------SLLCFAPDLIINEQRMS----------LPWMYD
44|GCR_XENLA/1    ----MLGGRQVVSAVRWAKAIPGFR--NLHLDDQMTLLQYSWMFLMV-FA-------LGWRSYKQTNG------------------SILYFAPDLVITEDRMH----------LPFMQE
43|GCR_ONCMY/1    ----RLGGQQVVSAVKWAKSLPGFR--NLHLDDQMTLLQCSWLFLMS-FG-------LGWRSYQQCNG------------------GMLCFAPDLVINDERMK----------LPYMTD
49|Q63449/1-16    ----QLGERQLLSVVKWSKSLPGFR--NLHIDDQITLIQYSWMSLMV-FG-------LGWRSYKHVSG------------------QMLYFAPDLILNEQRMK----------ELSFYS
13|NR42_XENLA/    ----DLLSGSLEVIRKWAEKIQGFV--DLPKEDQDLLLESAFLELFI-LR-------LAYRSRPEEG-------------------KLIFCNGVVLHRTQCVRG--------FGEWID
17|NR42_RAT/1-    ----DLLTGSMEIIRGWAEKIPGFA--DLPKADQDLLFESAFLELFV-LR-------LAYRSNPVEG-------------------KLIFCNGVVLHRLQCVRG--------FGEWID
36|NR41_HUMAN/    ----DLLSGSLEVIRKWAEKIPGFA--ELSPADQDLLLESAFLELFI-LR-------LAYRSKPGEG-------------------KLIFCSGLVLHRLQCARG--------FGDWID
70|NR43_HUMAN/    ----NLLTASIDVSRSWAEKIPGFT--DLPKEDQTLLIESAFLELFV-LR-------LSIRSNTAED-------------------KFVFCNGLVLHRLQCLRG--------FGEWLD
AA05172|CAA051    ----QLLTSSVDVIKQFAEKIPGYF--DLLPEDQELLFQSASLELFV-LR-------LAYRARIDDT-------------------KLIFCNGTVLHRTQCLRS--------FGEWLN
Y9|Q9Z0Y9/1-16    ----ELAIVSVQEIVDFAKQLPGFL--QLSREDQIALLKTSAIEVML-LE-------TSRRYNPGSESIT------------------FLKDFSYNREDFAKA-----------GLQV-
94|Q62694/1-16    ----ELAIISVQEIVDFAKQVPGFL--QLGREDQIALLKASTIEIML-LE-------TARRYNHETECIT------------------FLKDFTYSKDDFHRA-----------GLQV-
31|ECR_LUCCU/1    ----EITILTVQLIVEFAKGLPAFT--KIPQEDQITLLKACSSEVMM-LR-------MARRYDHNSDSIF------------------FANNRSYTRDSYKMA-----------GMA--
82|ECR_CHITE/1    ----EVTILTVQLIVEFAKGLPAFI--KIPQEDQITLLKACSSEVMM-LR-------MARRYDHDSDSIL------------------FANNTAYTKQTYQLA-----------GME--
83|ECR_MANSE/1    ----EMTILTVQLIVEFAKGLPGFS--KISQSDQITLLKASSSEVMM-LR-------VARRYDAATDSVL------------------FANNQAYTRDNYRKA-----------GMS--
46|O76246/1-15    ----EMTILTVQLIVEFSKQLPGFA--TLQREDQITLLKACSSEVMM-LR-------AARRYDAKTDSIV------------------FGNNYPYTQASYALA-----------GLG--
35|O02035/1-15    ----EITILTVQLIVEFAKRLPGFD--KLLQEDQIALLKACSSEVMM-LR-------MARRYDVQSDSIL------------------FVNNQPYPRDSYNLA-----------GMG--
43|Q92943/1-15    ----EMATNHVQVLVEFTKKLPGFQ--TLDHEDQIALLKGSAVEAMF-LR-------SAEIFNKKLPS---------------------GHSDLLEERIRNS-----------GISD-
15|PXR_MOUSE/1    ----DVSTYMFKGVINFAKVISYFR--DLPIEDQISLLKGATFEMCI-LR-------FNTMFDTETGT-W------------------ECGRLAYCFEDPN-G-----------GFQK-
39|Q91839/1-16    ----DIVTYMIKGIISFAKMLPYFK--SLDIEDQIALLKGSVAEVSV-LR-------FNTVFNSDTNT-W------------------ECGPFTYDTEDMFLA-----------GFRQ-
94|NRI3_HUMAN/    ----DINTFMVLQVIKFTKDLPVFR--SLPIEDQISLLKGAAVEICH-LV-------LNTTFCLQTQN-F------------------LCGPLRYTIEDGARV-----------GFQV-
95|NRD2_HUMAN/    ----MSFTPAVKEVVEFAKRIPGFR--DLSQHDQVNLLKAGTFEVLM-VR-------FASLFDAKERTVT------------------FLSGKKYSVDDLH-------------SMGAG
93|NRD1_HUMAN/    ----MSFTPAVREVVEFAKHIPGFR--DLSQHDQVTLLKAGTFEVLM-VR-------FASLFNVKDQTVM------------------FLSRTTYSLQELG-------------AMGMG
45|E75_METEN/1    ----EALLPAIRGVVEFAKRLPGFQ--QLPQEDQVTLLKAGVFEVLL-VR-------LAGMFDARTNAML------------------CLNGQLVRREALHT------------SVNAR
81|PPAS_HUMAN/    ----CTTVETVRELTEFAKSIPSFS--SLFLNDQVTLLKYGVHEAIF-AM-------LASIVN-KDGLLV------------------ANGSGFVTREFLRSL-----------RKPFS
32|PPAR_XENLA/    ----CTSVETVTELTEFAKSIPGFT--ELDLNDQVTLLKYGVYEAMF-AM-------LASVMN-KDGMLV------------------AYGNGFITREFLKSL-----------RKPIG
52|PPAT_RABIT/    ----FRSVEAVQEITEYAKNIPGFV--SLDLNDQVTLLKYGVHEIIY-TM-------LASLMN-KDGVLI------------------SEGQGFMTREFLKSL-----------RKPFG
46|RORB_RAT/1-    ----IQITHAIQYVVEFAKRITGFM--ELCQNDQILLLKSGCLEVVL-VR-------MCRAFNPLNNTVL------------------F-EGKYGGMQMFKAL-----------GSDD-
41|P97741/1-15    ----IKITEAIQYVVEFAKRIDGFM--ELCQNDQIVLLKAGSLEVVF-IR-------MCRAFDSQNNTVY------------------F-DGKYASPDVFKSL-----------GCED-
43|O02643/1-16    ----DKLTTMIQSIIEFAKLMPGFM--KLTQDDQILLLKSGSFELAI-VR-------LSRLIDVNRETVL------------------YGDVVLPIRDCVHAR-----------DPRDM
AC23439|AAC234    ----EISTKCIIKTVEFAKQLPGFT--TLTIADQITLLKAACLDILI-LR-------ICTRYTPEQDTMT------------------FSDGLTLNRTQMHN-------------AGFG
AD37372|AAD373    ----NVMDITMRRLVKMAKRLGAFN--EISEAGKFSLLKGGMIEMLT-LR-------GVTVFNADKGVWQTP----------------VDGHSQISFNMFDKLRP---------DIKDT
AB04069|CAB040    ----QWMYFNLFTTVEYAKTFMFFH--RLDSRDKFILTKHITLSCMN-LH-------LSYISV-ARKS-----------EVLT-------NPDGTSFPSRTQ-----------------AHYT
98|O17898/1-15    ----QWFFYNLLTTVEYAKTFMFFH--RLSSRDKLILTRHVALACTN-LH-------ISYSSI-RRNL-----------EVII-------EPDGSEPMTFND-----------------THYS
83|P91983/1-15    ----CWILYDLMVSIEYAKTFMFFH--QLDSRDRLVLMRYVALAIMN-LH-------ISYFTV-AKKY-----------DTII-------HPDGKLAPMKKG-----------------MVYA
83|P91983_1/1-    ----HWLYFNMLATIEYAKTFMFFH--ELLPRDQLILIRYTILAL-------------------------------------------HPDGTLQMHRNQ-----------------EYYS
49|O01449/1-15    ----YWTHCNMLTTIEYFKTFKFFH--ELSSRDKFVLARHTLLLCQN-LH-------ISHYTV-SHNF-----------DSCL-------QPDGSMQPKQDE-----------------RHYP
48|O01448/1-15    ----IWTLCNLLTITEYIKTFECFH--KISSNDRLVLTRHIILMCKI-LH-------ESHYAV-SRKF-----------DNYL-------NPDGTQESK-TE-----------------MHHS
39|O61939/1-15    ----QWRFSNLMTAVEYTKTFSFFH--DLSFQDQINLLKHVVIGLAN-LN-------TTYYTL-KNKF-----------DDLR-------QPDGTQRPGVNE-----------------PHYC
29|Q17929/1-15    ----QWRFFNLLTSIEFLKTFPFFL--KLDTKDQLILMRHVVVGCSN-LH-------SAYYTL-KNKY-----------EDLR-------QPDGTQRPKKSE-----------------PQYS
31|O01931/1-16    ----NYVHTSYLASIESSKMFEFSS--QIDLESRILLMKHATLICSN-LM-------NAFFSINEMKS-----------DILR-------HPDGSMSGHMRKFDREN------DVMIDHVK
29|O01929/1-16    ----NEVHSSCLASIEYSKTFDFSS--AIDISSKIILLKNTALSCAN-LT-------NAYTTFRKLKS-----------DTLL-------YPDGSIYGPPRR--KNG------PLIEKQRS
03|Q21803/1-16    ----GYAHNGSLAMVEYAKMIDFFP--KISKETAIKLVKHGLFMCGS-LS-------FSRRSIQKYNS-----------DILR-------FTDGSVAGKPKT--NWN------GVMLDHRR
04|Q21804/1-16    ----GLAHNSFLAVVEFAKFLDFFP--KLDSETQLKVVRHGALMVRG-LM-------NTERSIRKFHS-----------DCVR-------RADGSIAGKPMR--NYN------GIWVEQRK
69|O61869/1-16    ----NWAFHCSLVTVDFMKKFPFVN--LIRFEDQKYLLKEFYIKLIV-LI-------SSKRSMSRGK------------ECMT-------FPDGSDVLPPASSEWHIS----KISPNLENK
83|O17683/1-16    ----TWAFHSCVTTIDFMKQFPFVN--LIRSTDQMILLKESYVKLGA-LI-------SATRAYSSKK------------QCIS-------FPDGTDCLP--KTQWTVP----KISPNLENR
60|O45560/1-16    ----DWFYLVAVSTSDFLRKFPFVY--QSSSRDQSFILQKNFVKVAS-LC-------EAFRYYLLGE------------KQLT-------EPDGSNILIED------------LGIELGER
AA21008|CAA210    ----SWYYITLVSTVYFMKQFPFVY--FARLQDRIAILRETFLKVAL-LS-------ASFRAYALGK------------EVIS-------FPDGSSIFPSKETERIV------KDLNMPEL
61|O01561/1-16    ----EWSFLTTYVSVEMFLNLDFMQ--NIDNEDKIVLLREFSLKTLV-LN-------SAMRAILNKT------------DRVM-------TPDGKDVYPDRLLQM--------FSLDFLNN
```

Fig. 20 continu d:

```
---EWSFQATYIAVENFLSLDFMN--SIENDDKIVLIRNFSSKSMQ-LD-------SAMRSLNAKV-----------DRVI--------TPDGNEVYSDLLLAM---------FSIDFLNS
---DWTLLGLYVTIEMFLSMEFMG--SIDNEDKIILIRNFSFKSVI-LN-------SAMRSYSSKV-----------DRVT--------TPDGIDVYPDFMLRM---------FSPEFVIG
---DWCFFGAYTAVKFLLGLGFMK--KLSTRDKSLLIANYSAKATL-LF-------SAIRTMRAKN-----------DKMI--------KPDGKDFFVEFLSQWS-------DFSLHYTNR
---DWTFLALYTTVDLLLNLDFME--KLSTPDKLILIRHSASKCAL-LG-------GAMRTYLDKK-----------DRMT--------TVDGQDIYPKEMRALLGFQ---QGADQFLDR
---DWSFFALYTTVEFLSNMPFMS--KLQTAEKMCLLRNSASKCSL-FA-------GAMRSFNEKK-----------AEMI--------TVDGQDIYPNSILDFVGSS---GGAREFLGR
---EWSFLAAYNIIEYFTSFHFLD--KLNNDDKMKLFQANSFRSNI-FC-------GGMRGYKEKR-----------DKMV--------TPGGQEILPEILVKLLN------TSPQLLRR
---EWSFLAVFNVIEYFSTFDFSE--NLEYSDKKLLFQCNTFNANV-LC-------GAMRGFKEKR-----------SKMV--------TPTGQAILPDGIAPLLN------TTSQMLHR
---EWTFLGSYYKIAQFLDFDFVK--TMSLADRKILFSYNTLRMGS-LS-------RSMRTYLDKR-----------DCLM--------TPSGDDLFSPAVRGLFD------KTPEIINR
---EWAFLDVYSRISHFSNFEFMN--NISFADRKLLFSFNCLRTGV-FH-------GSMRTLREQR-----------DCLL--------TPSGEDVYPDAVHNLFK------DSPGLLNR
---EWSFYCAHSKILFFLGFDFIN--EMSTSDKKILFENNILRTGS-LS-------GAMHAYKANK-----------EIMM--------TPSGEDYVSEKIHQLFV------KSPNIITQ
---QWSIMMAYSRISYFLDFNFIR--DLDSSDKNILFKYNISRAGC-LA-------LAKCAFNENK-----------PKLT--------FPNDVDAFPSEMYNLCG------SSTAVLNQ
---EWQIMNKLSTIEMLKNLDFPK--YLSSQDLRYFLKKGYFLKGI-LT-------MAMRSYLNQE-----------EFMS--------FPNGVDLFPDHAREGLH------INISFLNL
---DWGFMNQVTIIDFLKKLEFSK--FLSSQDLRSFLKCTHFTQVI-LK-------SAVISYNANQ-----------NYMS--------FPNKIDIFPETVTEGTS------ISVNLQNR
---PWGSLGVLLAVEVCKGLSFYN--DLLLSDRVLIIKNVAFKSHH-LS-------VAYDSFIQKK-----------GRVL--------APTGTEMFPDVLFEIP-------KCREIIMD
---YWHERDWFVMIEWAKTLPVFQ--SLPFTDKLALIRHSAITYPS-LV-------HVFNSP-DHGL-----------DTIV--------FPDGAYFDRTPEPTRPLG------FNKKKYQ
---KVLKWSINNAIEWASQFSPFQ--LIMNEDQKCVISEYGFAFLV-VD-------QAFRTISETPN-----------GCWY--------LQNGTYLNSDYLYGLTDE (8) IRKHSEFVKS
---VMLKLCLKQATEWANRLKLFK--SLPAESKRSILAEYCLAFNL-ID-------QGYKTSKEADL-----------GIWL--------LQNGSFMHPDYFFGLNVT (11) QLHHNFVAE
---LMLRIGFMNATEWGNCFRPFR--YSNRSVKKRILSEFGIASGL-ID-------QAFKTATETNQ-----------NEWL--------FINNSCLRPDAFC---------T-VHNDFLNT
---EITSVQNKLAAELCKTCPGVD--LLDNEDILIINKYFQFSNVW-IE-------STWNYLR-E-------NNSVPIDDSE-------IDLKLLKFINQVK------------------
---EITSIHNKLAAEICKTCPGVD--LLDNEDILTIAKYFQFSNVW-IE-------SAWNYLMSA-------DKHVQIDDSE-------IDLTLLKFINQVK------------------
---EIVSYQSKVAAETCRTCPGVD--LLDNRDILIILKYFQFSNIW-IE-------STWKYWF---------SNDFSNDTEK-------FDIKLMEFIGQVK------------------
---KIVATQQELALEACRTCPGMD--LLDKEDRKVVQKYFVFSNVW-IE-------STWLYSLAK-------EHLETENNLN-------FDINLKKFIEQVK------------------
---RLVNFQRKIADEICSSFPGVE--LIDENNILLILKYLQFSSVW-IE-------STWTYWIT--------NSSVSVDNME-------MDKRLKELIEHVK------------------
--NELVTLQ-RTSTDSCLHCPGVD--LLDRESAQALLKHFQFADVW-LD-------SAWEFSKSPELLFAL (7) LNSLNNMN-----NTNQELSEFISLVK------------------
---LITAHERGSAMEACIDCPGVD--HLEQEDLKVLLKYFQFANVW-ID-------SVRAYSVSDID---------CFKEVP------GRDDRLTEFLQQVK------------------
---VITAHQSGSAMEACINCPGVS--KLEQKDIRVLEKYFQFLNVW-LD-------SVREYSMSDG----------AHYEI-------GVDRKLSEFIHQVR------------------
---LISAYQNGLSMESCMHCPGVN--SLEKEDVQTLLKYFQFANVW-MD-------SVRAYSMSNV----------DIYETT------PKDKRLSEYIQQVK------------------
---LISAHLTGSAMESCFQCPGEN--LLEKEDIQVFLRYFEFSNVW-ID-------SLREYSLMSHS---------HNKEIS------RTDKKVMKFINQVK------------------
---LITARQSGPSMEECLRCPGVD--MLEKEDVEALLRYYRFSNIW-VD-------SVLIDSVLESA---------KGDSHT------DTDDRLSRFIQEVK------------------
----VSAHHEGSSTESCLNCPGVD--MLDKEDVEVLMKYFKFSNTW-ID-------SIFASLISPN----------QVELIQ------CDHEDIAKFINHSK------------------
---KISEMENNVVLDDCLNCPGVD--VLEQDDVEVLLKYFKFVNIW-VY-------SALASSSTCLE---------QFPEQTG-----DNSNQLFDFLHQVR------------------
---DLSDVLNDLSMESCMNCPGVD--ILRQEDIEVLLKTFHFANVW-ID-------AVWAFSRHCES---------ICDSTF------EDNEPLYQFINQVK------------------
---DISETLNDLSMEACMNCPGVD--ILRQRDVEVIIKSFHFANVW-MD-------SIWAYSRHNES---------IVNKFF------ENNGPLSMFINQIK------------------
---GEVKHDFVMILDYVNQFPEFQ--ALHKNDKNVVFRMACAVDSM-LA-------SAYYS--YKVGIEKE (7) DYINMNP (12) EFQIPQEHEKYKTLMP---------------LKLK
---QKNRNDYTMILDYINRFPEFE--SLSKSEKTVLFRTAAAVDVL-LD-------QSYYSQVYVIFPTED (7) EYLPMNP (11) NFHSDEDYDRFKSVKNPVKLC (7) -RMLTSMKVR
---DIIEYYFTTVIKWISHFDEFR--KLDQNIQMKLVQSIWHILMK-FHKCSA---TVAYRKSSDEKTRQLQKVIRNVCLD----REKMKMDFSWMSDYPAHHVL-----------RCMF
---DIIEFYFMAGVKWVSHFDHFQ--KLDQDLQMKLAQSIWNIFIK-IHKCSN---TAKYRKSNSIGTHD-QKILRNMCVD----REKVKFDFSWMSDYPQHFVK-----------KYMF
---DIIEYYFLTVAKWIAHFDEFR--KLDQSLQIKLLQAIWHVWSK-THKCAS---TAFYRKSNPNAKPT-QKILRNIMQS (12) FEKWRFISAKSSKFHRNRCKNCI------HFRFML
---EMWEYYFITVTKWLMYFDEFQ--KLNRQIQMTMLQAIWHVFSK-LHKYVA---TSAYQKATPRAKPT-EVIIKDVMMD----IETVNFDCSWMSDYPTKDVF-----------KFLS
---DTWQWQFLTVAKWLTHFKQFD--ALEEKVKLTLLQSTWHVWSV-LDHHFA---SAAHHRSNPNALKTHAVTRRGVMMD----MTNVHFDANWLSGYPSAELN-------------RF
---KIWEFYFLTVARWLTYFDEFQ--KLDHKIQMTLLLAIWHVWGR-IDKLLA---TAINRRRGLCPTKNLLTLSNGVFID----MEHQEVDVDWMTNYEPEQVLT----------FIDG
---DCWEYYFLTTATWLTNFDEFQ--KLEHPVKMKLIFAIWHMWGR-MDKLAC---TALARRRQQLSAKSEIAQSNGLFLD----VDTLALDVSWISEFKLEEVEF----------FIDG
---YFWEFYFLRTAKWLTYFDEFQ--RIPDEIKIKLLLSFWHVFAR-LDKLIT---TAKARKLKLCS-QQTWAMSNGLILD----FDRTKVDFSDISNYPTEDLFY----------FLNS
---GFWEYHMLASAKWLTYFDEFQ--QLDQDVKRKILFGVWHVWGR-IDKLMG---TALFRIKNKEASLSDRVVGNGIVTD----MNCVVAETKWMSRLPVQQLRY----------FIDG
---TFWETDFLTVAKWLTYFDDFQ--LLPHNQQILLLKSVWHVWNR-LEKLAL---TATSRRQNVCEQKQLMLTYN---SVCN--PKTIEELDYSWFTKYPKEQLCF----------FFDE
---GLWQDDMMKVAKWLTYFDDFQ--QLSHRMRSEMLKGMWKVWSR-LETLAL---TAMGRRLMICQKDMIMTHTEKEQVMAH--PKQIEVDISWCSRYTLEQLRF----------FADS
---SFFEQDFLKATRWFTYFDEFQ--DLDQEQKLELIQAIWHVWSR-LYKLST---SAIGKRRQICEDKMLMISHGSEYTVMD--LTKIEFDFSWCTKYTNEQMQH----------FIDN
---LIWEQEMLSTAEWLTNFQEFN--SLPTNVQLEVIKSIWQTYGR-FEELAR---TAEFRRKKLFKNNNVFMVGDEACLD----TENVEVDVSWFSDYTYEQV---------SYFMDC
```

Fig. 20 continued:

```
77|O16677/1-18   ----LIWEQKILNTIDWLTNFNEEN--VLPLDVQLKIVKTIWQVYGR-FEKIAK---TAELRRKKLIEHDNVFMVGDEACLN----SKNVEMDVSWYTDYSFEQI---------AYFMDC
59|O16359/1-18   ----KTWEQGFLRAVEWFSNFSEER--ELDENLKMEIIKTCWVSWIR-IDKLSE---TANKRVNATLD-NSLLMVGNDSCMH----MNDYEVDLSWCTNYSLEQLA--------FFFLTP
60|O16360/1-18   ----YTWEQGFLRVVEWFSYFPEER--ELDESVKLEIVKTSWVAWTR-LDKIFE---TAANQKNGVFG-DSILMVGNECCMD----MNDYEVDLSWCTNYSLEQLA--------YFFLTP
61|O16361/1-18   ----LIWEQAFLRAVEWFSNFSEEN--AIDELSKMTIIKSAWISWTR-IEKLAE---TADHQRKKIFG-DSVMMCGNDACLD----LANYEVDLTWCTNYTTEQL---------LFYLSP
58|O16358/1-18   ----YTWEQDFLRVIKWFANLDEEN--DLDEDLMLNIVKAAWIPWTI-IEKLYE---TSDYQRKNIFS-KTVLMCGNDTCMD----MNNYELDLSWLTDYSLDQLT--------YFFTPP
55|O16355/1-18   ----VEFEQMFLATANWYARLPDET--SLDLRVKLETAKSTWMLWIR-INKLAE---TAQNQRIKEMG-ANIFMCSENTCLN----VQDINVDWSWCTNYGAEQM---------QAFLMP
93|O44593/1-18   ----RFWENSFLKTAQWESTFAEER--ELDTKIKTDILKSAGGLWIR-IERMSE---TSVQRRKQNFG-TNVMMCGEGKCVN----VNEFDMDLSFCTHHSMDQL---------KPFLLA
66|O16966/1-18   ----MVWEFDFISTAKWLTRLQDEC--TLPMRIQMQLLQTIWHVWSR-MYKVVK---SSELRKKHADS-SNLFQVLDQFHVN----LGITKLDVSWLTTYSYEEIKY---------FLYG
38|O44638/1-18   -----DFQTYMCAAAKWLSYSDRER--NLDDELKIQILQCIWFNWGR-IERIAT---TAKMRNKRKCG-KKQFVFSNNILLD----YDEMSSDLSFWSNYTFEEMK-----------YFF
37|O44637/1-18   ----KQFEKAMCATAKWLACSKKIQ--SFDDDVKIKILRTVWFNWGR-IEHMTA---TAKMRIRGSCG-KEQLMFSSDTKLN----FGNVHIDLSCFSKYSFEQMR-----------YFF
15|O45315/1-22   ----QWWEVQMRDTATWLMHSDEER--RLPSHEKIATEKIVWAVWRR-FERYTM---SAQMFGQKCYDEKIHVHNHQFAARF----G-NYHIDYSNVSDHGFERFNN(26)KFNFFSVFG
60|O45460/1-19   ----EFWESQMEDTATWLMYSGEER--KLPNHEKIATEKIVWAVWRR-IERYTM---TAQVFGQKCYDEQILLHSHQDAARF----T-DYDVDYSYITDQGFEKIN----------GLFG
54|O45954/1-21   ----SYWKIHMEYVATWLMHSNEER--SLPNHEKLATEKTVWAIWRR-FERYAM---SAEIFGKRCYEEKTKFFEFSTFLRS----I-ERRETRHSRCRRLFSLLAR(23)AKFWFDVLG
48|O18048/1-18   ----VIFEKQMVMIAEWLIHNPDER--MLDEEEKYLYEKAVWNMWRR-FERFEM---SVKMFGNQVL-TKRKFAISNEKLMV----M-DSCIDYSEITDLPNSKVA----------EMFR
35|O44635/1-18   ----VFCEQQMVIVATWLAHCPYEQ--KLESTLKYQMEKVVWNVWRR-FERVEM---SVKMYERETPGEVMKFAISDKYYVT----P-KLKIDISKVTNWPVPFMG----------GIFR
87|O18087/1-20   ----DVQFKVIGSCATWLLFSSFEQ--KLEENEKVVILERLWHGWTV-LEFLSR---SLEIFGNKVIDEKVKLILIQQLKIF----F-SNIRIVFISDNTAIRLITA(12)SESIKKKLE
10|O45910/1-27   ----QWWAEQLEDTATWLMNCREER--GLPIHEKLATEKIVWAVWRR-FERYT(29)CVDYSHVWSQGPHTF(7)MIRYFDI(12)TEVIYILCQIVWNYGFRPLSG(32)SKFLTEKFN
06|Q15406/1-16   ----RLADELLFRQIAWIKKLPFFC--ELSIKDYTCLLSSTWQELLL-LSS-------LTVYSKQIFG--------------------ELADVTAKYSPSDEELHRFS------DEGMEVIE
99|O77099/1-16   ----DSMKQQLLILVEWAKSIPAFA--ELQLDDQVALLRAHAGEHLL-LG--------LSRRSMHLEEMLLLG--------------------NNCIITKQSPDSKMAPN-------LDISR
82|P97782/1-16   ----RMADQTFISIVDWARRCMVFK--ELEVADQMTLLQNCWSELLV-LD--------HIYRQVQYGKEDS------------------ILLVSGQEVELSTVAVEAG------SLLHSLVL
32|Q9YH32/1-17   ----NLADKELVHMISWAKKIPGFV--ELSLVDQVHLLECCWLEVLM-IG--------LMWRSVDHPG---------------------KLIFCPDLSLSREEGS---------CVQGFVE
33|PPAS_XENLA/   ----CTSVETVRALTDFAKRIPGFG--TLYLNDQVTLLKYGVHEALF-CM--------LASLMNKDGLLVAG-------------------GRGFVTREFLRSLRQP-------FCHIME
27|NRI3_MOUSE/   ----DINTFMVQQIIKFTKDLPLFR--SLTMEDQISLLKGAAVEILH-IS--------LNTTFCLQTEN--------------------FFCGPLCYKMEDAVHAG--------FQYEFLE
92|FTFB_DROME/   ----NVADHRLYKIVKWCKSLPLFK--NISIDDQICLLINSWCELLL-FS--------CCFRSIDTPG---------------------EIKMSQGRKITLSQAKSNG---------LQTCIE
60|O16360_1/1-   ----KSWEQGFLRVVEWFSNFSEER--ELNERLKLEIVKSCWFSWTR-IDKLS---ETANKQINKMLGKSQ-LMVGNGACMN----MNNFEIDLSWCTNYSLEQLK--------YFFQTP
03|DAX1_RAT/1-   ----EAASAGLLKTLRFVKYLPCFQ--ILPLDQQLVLVRSCWAPLLM-LE--------LAQDHLHFEMMEISE-----------------PNLMHEMLTTRRQETEGPEP(5)TEQPQTVSAE
96|Q22296/1-16   ----GWAIMNKLSALEKLRSLDFPK--YLSSKDLKYFEKGAYFLEAM-LT--------MAMRCYFCKEDLMH--------------------FPHGVDLFPVSVQEAG------KIKLNFLNL
FTF1_BOMMO/1-1   ----KVLDQNLFSQVDWARNIVFFK--YDKVDDQMKLLQDSWSVMLV-LDHLHQR---MHNGLPDETT--------------------LHNGQKFDLLCLGLLG---------VPSLAD
AB03298|CAB032   ----AIFEKQMIVVAEWLVQNPDFR--LLSLEERYIYEKAMWNMWRR-FER---FEMSVKMFGNDVVEIRKF(3)-NDQILT-----KEFTIDFSEITDVPNTTVQ--------EMFRHS
34|O17934/1-18   ---IVFQNESTSIYQFIWEAFPEXS--MLLPDTKLAFEKNFELSFCL-FESS----FNGCSAKQKNVMLIPS-----G-------DYIDLAHSESFFINDHQTFTER-----DNIEILAQ
36|O44336/1-15   ----EITILTVQLIVEFSKRVPGFD--TLAREDQITLLKACSSEVMM-LR--------GARKYDVKTDS--------------------IVFANNQPYTRDNYRSAS---------VGDSAD
96|HR3_DROME/1   ----EKLTQMIQNIIEFAKLIPGFM--RLSQDDQILLLKTGSFELAI-VR--------MSRLLDLSQN--------------------AVLYGDVMLPQEAFYTSDS--------EEMRLVS
42|O45842/1-17   ----CWLFLTAVSTIDFMKKFPFVF--LSNSHDQTILLRKKFVKVAS-MC--------EAFRVYLFRDK--------------------QLAFPDGSSILIEGLDSELG------------DR
AA18369|CAA183   ----YFGITSCAACAAFFRRSDGIRY-MCTATNSCTISHDMKFFCRA-CRYTS(29)PRSVPSLEEVDGFSK(15)EVESIM(12)QVKSVNDLMIVTAYQEGTSLE(34)APTSPLDPS
AD38900|AAD389   ----KIGDSIVHKMVAWTKKLPFIM--EIPMEIHSKLLMEKWHEISV-LTT-------AAYQAMHGKQTHAPP------------------SSDHEQDFMQEVNANLRTL(5)SLMGRPITLE
80|P91780/1-15   ----DISTKSIIKTVEFAKSVPGFT--DLTIADQITLLKSACLEILF-FR--------ICSRYDNSNDTMS-------------------FSDGLTLNRDQLRNCAFG---------PMTE
73|O61873/1-16   ----HWAFHCCLVTVDHMKKFPFVN--LLRYDDQKYIIKKYFIRLKA-TIG-------AHRSMMSGRD------------------RLIFPDGSDVLAQAYSEWDLT----TIPSHLIHK
74|ERR1_HUMAN/   ----DLFDREIVVTISWAKSIPGFS--SLSLSDQMSVLQSVWMEVLV-LG--------VAQRSLPLQD--------------------ELAFAEDLVLDEEGAR-----------AAGLGE
AA16413|CAA164   ---IVSSKECDLVADWILNSYPGFA--DLHKDQKIILERNFFLPFVI-LQGG-----HFACIHNRSDVIILAS------------------GDYIDCSHPETFYYDPDG-----RQLMSSED
76|7UP2_DROME/   ----EIAARLLFSAVEWAKNIPFFP--ELQVTDQVALLRLVWSELFV-LN--------ASQCSMPLHVAP-------------------LLAAAGLHASPMAADR----------VVAFMD
77|O44577/1-18   ----YKNRNDYIMILDYVNRFPEEH--LLNKSEKTVLLQTTAAVDAL-AD--------PAYYSQVIYPDEP-------------------VFVTREAKYLKMDPMPSTELE(7)-LEDITIYKN
70|HR38_BOMMO/   ----SILTTSIDMIKLFAEKVPGYG--DLCPEDREQLEASARLELFV-LR--------LAYRTALEDT--------------------KLTFSNGSVLDKRQCQRS--------FGDWLH
03|O61203/1-18   ----QDMERSMFGAAKWMMCSERIR--NYQDSLKIRLIQAIWYTWGR-IERIAT---TAKLRLEKKCEEKH-FVFSPMFALN----LDEMSFDASCWTTLPFEEMQ---------YFYVP
99|MCR_RAT/1-1   ----RLAAKQMIQVVKWAKVLPGFK--NLPLEDQITLIQYSWMCLSS-FA--------LSWRSYKHTNS--------------------QLLYFAPDLVFNEEKMHQS----------AMYE
AA18351|CAA183   ----YWHERDWFVLTEYAKTFKVFK--NMPYHDKTELVCHAAIVIPV-LN--------QVYNSPDYGLDT-------------------VVFPDGTYYDRTHEPTRPAG----------LNRK
05|Q21805/1-12   ------------------------------------------------------------------------------------------------------------------GFAH
93|Q18393/1-18   ---KLSKMDSKLLWDLGEVIFSDVS--KFNSFDKDSMEANFFPKWII-MECA-----IDYSLNQEQFQSFIG-------------------SIEYYKKCAHFYGSSMPE-----EKRLDDED
07|O45907/1-18   ----SFFETDFLCATKWFTYLDEEQ--FLDKNQQLTLMQGIWHVWSR-IHKLAV---SAMGRRRGICDKNTV(5)NEFAVCDLN----EIEVDMSWCTNFSNEQMR---------YFLDT
```

Fig. 20 continued:

```
64|O16664/1-18  ----------YVGRPEFLIFFDDSNVTTQKTFIDIITVQAIWHVWSK-VHKCAS---TALYRKSNPNAKPEQ-KIFRNMCMDR----NLGKFDTSWMSDYSTEHVIR--------FMLTP
09|O16609/1-16  ----PWGSLGMLLAVKTIKSLDFYP--KLLLSDRILLLETVTLKSFH-LS-------VAFDSFSQKKE------------------KILAPDGSEMFPDSLNKMD-------FCKDLIMR
54|O16354/1-18  ----LFLEQSFLGAAQWFARLPEFS--MLDPQIKIDILKTSWMIWAR-LDKLAA---TADFHRQKLLGNDVY(5)-TCMN------PGNVKIDIKWSSNYSVEQLR---------GFLVP
46|O01446/1-15  ----IWTLCNLLTTVEYLKTFNYFK--DLSALDKLILARHVILIGKN-LH-------ISNYSFSKKFTG-------------------CLNIETDAQRERHYP-----------------
17|O45117/1-17  ----AMVNKQVFQHMDYCRRLPHFT--KLPLNDQMYILKQSLNELLI-LN-------IAYMSIQYVEP-------------------DRRNADGSLERRQISQQMCLSRN--YTLGRNMA
55|E75C_DROME/  ----QRFAHVIRGVIDFAGMIPGFQ--LLTQDDKFTLLKAGLFDALF-VR-------LICMFDSSINS-------------------IICLNGQVMRRDAIQNGAN--------ARFLVD
AB07282|CAB072  ---CSVSKDIDLTWKMTYKLFPSTG--KLKKSDKIALLRNFIPKLMQ-HDP-----LLNYIRNFKKYEAME-----------------QPDLEQIIINFYKGAMPE(23)TTKITPLFL
27|O61227/1-16  ----ESGSRLLFLSIRWTRNIPAFQ--YLTTKTQITLLRGCWAELFT-LG-------LAQCSQTLSLS-------------------TILSALISHLHTLIAQDKMS---ATKVKQVSD
49|RORG_HUMAN/  ----HHLTEAIQYVVEFAKRLSGFM--ELCQNDQIVLLKAGAMEVVL-VR-------MCRAYNADNRT-------------------VFFEGKYGGMELFRALG---------CSELIS
65|O16965/1-18  ----KVWEFDFITAARCIANLDGFS--ELPDKLKMQILQSIWHVWAR-LFKAAR---SAQLRKRQGLFSNAL(5)-------LYVEMETIRMDVTWLTRYSFDQVK---------YFLYG
26|O17026/1-18  ----AFWKFLNKRYKSAQIFFAFFP-----N-FQLKILECIWHIWAT-LDRHS---ATAVYRKSYPNAAPTQIVSRRGVLMD----LAKINFDSTWISDYPAVN---------IVYFLRQ
46|O16346/1-17  ----FVLKLGLENANEWGNQFEPYR--RLSQTDKNSVLSEFGFAFLL-LD-------QGYKTAQRADEG------------------FWLLQNETFMHPNYFFALSVED----AMKENAEQ
13|O17013/1-17  -------MWSFNDAVEWASDYEPFQ--LLPNELQKCVISEFGVAFFL-LD-------QGFKSARESKRG------------------FWMFQNGTFLHFDYFHGLAN------IDDTADRE
64|O45664/1-21  ----DWAFLGIHARISYFMDFEFIN--KLSLNDKVILLKKYNALRMGC-MT-------EAMRCFTAKRAY-------------------LVTPSGEDVYLPIFRRIFEK------TPALLDQ
80|O96680/1-15  ----ETTARLLFMAVRWVKCLMPFQ--TLSKNDQHLLLQESWKELFL-LN-------LAQWTIPLDLTP-------------------ILESPLIRERVLQDEAT----------QTEMK
42|O61942/1-16  ----DWRFFTMYTTIEFLLNLDFME--KLEHSDKMVLLQNFSAKATL-LFN-------SLRTVRGNNDR-------------------LTTPGGHEIVPDLMSTLFN------VSLSFIRQ
30|O01930/1-16  ----CYIHCGYLASIEYSKMFDFAH--KIDVASKATLIKHATIMCAD-IM-------TAFFSYYQRKSDR------------------LIHPNGMFAGPPKYRYGEAG----------TKYQA
00|Q21700/1-18  ---EILSNEMSLVADWIGWCENEFE--FLPTDQKSILFSNFSVNFTV-LER-------AFLTSKHNNLN-------------------QWVIQSGDYIQIDNLEEFFT-(6)DGTKLAKLFK
57|O16357/1-18  ----LFWEQDFLRTARWLSHFDEFQ--ELPLSIRMQILKVAWILWGR-LDKIVK---TANERKNQQLNQNFYMIG-KDVGIDIG----NEEVDVSWCTNYSKEQLA---------YYLDC
88|O16388/1-16  ----EWSFFGIYTSIEFLMSLDFMSSLNLNIQDKKILLKHYCVKASL-MA-------SAMRAMKEQRN-------------------TLMTVDGKEIYPDAILRLKQIS------PQFLQR
87|O45987/1-16  ----FVMSLGFQNALSWANHFETFT--KLSEDQKALVMSEFGIAFIL-LE-------QAFKTAEDTEEG------------------CWILQNDTFLEISDTRNVQG-------QKRINSE
86|O18086/1-18  ----DFRLGHIKNCATWLTHSSFFQ--SLTETENLLILKSTWHVWSW-LELLS---VSVEIFGNQVCEEKIVFLS------------EKIAVDIVKVFRYILKPLN---------KQEKRK
32|O17932/1-19  ----EWSLMLSFSKISYFMSFTFIN--ELSLADKLIILFTANILRTGT-MA-------RAMKSVLEKRS-------------------ILIFPNEENIFPVEMNIMFE------NSDVLVEQ
AD03686|AAD036  ----GMSKHDFVMIMDYARGMPGFD--EMNYADSVFCYRLVCAVDFV-IN-------SAYYTYKRGIEHNEL(5)----------TFIPMVPTPLTGYEENANLLFQS-----QDDLMKFR
FXR-B           (20)ETAVLHIRGLMNFTKGLPGFE--NLANEDQTALQKGSKTEVIF-IH-------GAQLYNTMIIS-------------------ICLILPYVWMKIHFRISFLG-------VTEEFIT
20|O97120/1-16  ----QSIEQQLPRIVNWARQLPVFSSVYLSFDDQFCLLKAAWPELVL-LS-------SAYHSTVIRDG-------------------LLLSIGRHLGREVAKSHG---------LGPLVD
16|Q20916/1-18  ----EWAFLATYSHISYFLNFDFVN--ELHDIDKVCKLEKLPHIVDV-FKKCI(12)GVMRTVNEKRS----------------K---MMNPGGQEIYPDVILQLYE------KSKSTLHR
31|O17131/1-18  ----TNWEYYFKRVATCLNHFDEFK--KLEIPMQLKLLQTIWHVWGR-LEKITS---SARYQKTVENSKLS-(4)QSGWIVD----VKDVDIDCNWFSNFPNEQVKN---------FMLQ
AD34463|AAD344  ----NIMDVTMRRFVKVAKGVPAFR--EVSQEGKFSLLKGGMIEMLT-VR-------GVTRYDASTNSFKTP-----------------TIKGQNVSVNVDDMFAKLN---------ANAQAQ
16|Q93516/1-16  ----EWGFLDSYSRINFFLCFDFIN--NLSLVDKNILFNSNTMRAVL-LG-------AAMNTKCHGVDILR-------------------APDGADIYPDEAIKLFP------TDSKILDR
34|O62034/1-17  ----QIQSLSNLDKADFSKLSKFAF------FLFQNAVIRNFSPKWSL-LASH-----IDFERNCHLYEDVVT------------------VDDYLDMIVKFYSGSMYG-----DNKISPKE
68|O61868/1-16  ----NWAFHCCLVTVDFMKKFSFVN--LLRFEDQKNLMKEFYIKLTV-LIN------SKQSMSCGKEG-------------------MTFPDGSDVLPPTSSEWGIS----KISQNLENK
91|O17991/1-15  ----FVMSLGFQNATVWANHFEPFW--SLSNDEKTSVLSEYGIAFTL-ME-------QAIKTVNEG-----------------------IKGVWLLQTGTS-------------------
53|VDR_RAT/1-1  ----DLVSYSIQKVIGFAKMIPGFR--DLTSDDQIVLLKSSAIEVIM-LR-------SNQSFTMDDMS-------------------WDCGSQDYKYDVTDVSKAG------HTLELIE
AD33899|AAD338  ----KIMTPAITRVVDFAKKLPMFS--ELPCEDQIILLKGCCMEIMS-LR-------AAVRYDPESET-------------------LTLNGEMAVKREQLKNGG---------LGVVSD
64|O16964/1-18  ----HFWELDFLKTAKWLSFLDGFY--SLPRGVQMQILMSTWHLRAR-LHRLS---QTAKLRKSEDIGKND-FMMSKNSCLD----LKTCQLDVSWCIDYPNEQIQ---------FFLEN
35|O02235/1-15  ----EISKSEDKTSLDAVLHCYGTY--ILDDEDITTLLRYFKFMNTW-ID-------SAFVYSKSTSN-------------------EELLDGNDICKFAYQID----------------
06|Q21806/1-16  ----TYMHSALLAAVEASKTFEFFS--KLSHEARMILIRHVSLIGSN-MM-------SASFSMHHRKS---------------DE--LLLPDGTVFGSIGGCLASEVLG------EIKYK
54|O16754/1-17  ----SSWEDQILAVVKWLNYSNAFR--NLPYHLRLQLVQTTWSTWNK-LER---TAMTAQMRVNRQCGKN-------------------QFVVSNNHLSDYSRMDAD---------MSCLFN
56|O45756/1-18  ----KKWENDLIRIANWLLHSEHFR--MLPLEEKMMILKKINARWRK-HEKNL(4)TFGHKVYKENVIVCIN-----YNAAN----LNKVHIDYSNITDRSNEEMS---------QLFGRQ
29|P91829/1-16  ----KSIHQQLVLFIEWAKSLPQFS--FLAQADQAALLKGGAASITV-LG-------VAYRSICLTVEN-------------------TICLANDTLLPKEHATQVG--------DINCVVG
NH64_CAEEL/1-1  ----DSMNQQLTLMVEWAKVLEGFQ--RVDNITQVALLRHFSACHLV-MC-------AAFRSIHLSDA--------------------VWLTNETCLHKDSPKIP----------DMNRVAE
```

L-0020-01-WO-01.ST25
SEQUENCE LISTING

<110> LION Bioscience AG

<120> NOVEL MAMMALIAN NUCLEAR RECEPTOR L66 AND METHODS OF USE

<130> L-0020-01-WO-01

<160> 24

<170> PatentIn version 3.0

<210> 1
<211> 1152
<212> DNA
<213> Homo sapiens

<220>
<221> n
<222> (1)..(1152)
<223> n may be a, c, g or t

<400> 1
cctggaataa aaaggtccag accaacctat tcttcctcga gaaataaggg acaggaagaa 60

ttctgtgtag tttgtggtga taaagcatca ccatcaccat atcattataa tgcacttacc 120

tgtgaaggtt gcaaagaaat acctatggta aaaaatttta aaactttttt attgggtttt 180

tttcaatgta gcatcnnnca aaatgcagta tatagttgca ggaatggtag tcactgtgaa 240

atggacatgt acatgcgtag aaaatgtcaa gagtgcagac tgaaaaagta taaggcagta 300

ggaatgttgg cagaatgttt gctcacagaa atccaatgta aattaaagag acttcaaaag 360

aactttaagg agaagaatca tttttactct aacatcaaag tggaagagga aggagtagac 420

cacagttttc tatcatccac cactagacct ggaaaagaaa gcatggaact aactgaagag 480

gaacatcagc tcattaataa cattgtggct gctcatcaaa aatataccat tcctttagaa 540

gaaacaaatt tgtatctgca ggaacataca aatcctgaac tgagcttttt gcaactctca 600

gagacagcag tcctacacat acgtgggcta atgaatttta ccaaggggct cccaggattt 660

gaaaatttgg ccaatgagga tcaaactgca ctacagaagg gatcaaaaac tgaagtgata 720

L-0020-01-WO-01.ST25

tttctccatg gggcccaact ttacaataca atgataattt ccatatgttt gattctaccc 780

tatgtttgga tgaaaataca ttttcgtatc agttttttgg gtgttactga agaatttatt 840

acannnctgt tttacttcta caaaagaatg agcaaacttg atgtaactaa tactgaatat 900

gctctgcttg cagcaacaat tgtttttttca gatcgtccat gccttaaaaa taagcaatat 960

atggaaaatt tannngaacc agttttacaa atattgtata agtattcaaa aatgtatcat 1020

ccagaagacc cannncattt tgcccatctc atatggaagc atactgaact gagaactctg 1080

aattataacc attcagaaat acttagcact tggaaaacaa aggaccccaa attggctact 1140

ttactctctg ag 1152

<210> 2
<211> 1152
<212> DNA
<213> Homo sapiens

<220>
<221> n
<222> (1)..(1152)
<223> n may be a, c, g or t

<400> 2
ctcagagagt aaagtagcca atttggggtc ctttgttttc caagtgctaa gtatttctga 60

atggttataa ttcagagttc tcagttcagt atgcttccat atgagatggg caaaatgnnn 120

tgggtcttct ggatgataca tttttgaata cttatacaat atttgtaaaa ctggttcnnn 180

taaattttcc atatattgct tattttttaag gcatggacga tctgaaaaaa caattgttgc 240

tgcaagcaga gcatattcag tattagttac atcaagtttg ctcattcttt tgtagaagta 300

aaacagnnnt gtaataaatt cttcagtaac acccaaaaaa ctgatacgaa aatgtatttt 360

catccaaaca tagggtagaa tcaaacatat ggaaattatc attgtattgt aaagttgggc 420

cccatggaga aatatcactt cagtttttga tcccttctgt agtgcagttt gatcctcatt 480

ggccaaattt tcaaatcctg ggagcccctt ggtaaaattc attagcccac gtatgtgtag 540

gactgctgtc tctgagagtt gcaaaaagct cagttcagga tttgtatgtt cctgcagata 600
caaatttgtt tcttctaaag gaatggtata tttttgatga gcagccacaa tgttattaat 660
gagctgatgt tcctcttcag ttagttccat gctttctttt ccaggtctag tggtggatga 720
tagaaaactg tggtctactc cttcctcttc cactttgatg ttagagtaaa aatgattctt 780
ctccttaaag ttcttttgaa gtctctttaa tttacattgg atttctgtga gcaaacattc 840
tgccaacatt cctactgcct tatacttttt cagtctgcac tcttgacatt ttctacgcat 900
gtacatgtcc atttcacagt gactaccatt cctgcaacta tatactgcat tttgnnngat 960
gctacattga aaaaaaccca ataaaaaagt tttaaaattt tttaccatag gtatttcttt 1020
gcaaccttca caggtaagtg cattataatg atatggtgat ggtgatgctt tatcaccaca 1080
aactacacag aattcttcct gtcccttatt tctcgaggaa gaataggttg gtctggacct 1140
ttttattcca gg 1152

<210> 3
<211> 384
<212> PRT
<213> Homo sapiens

<220>
<221> X
<222> (1)..(384)
<223> X may be any amino acid

<400> 3

Pro Gly Ile Lys Arg Ser Arg Pro Thr Tyr Ser Ser Ser Arg Asn Lys
1 5 10 15

Gly Gln Glu Glu Phe Cys Val Val Cys Gly Asp Lys Ala Ser Pro Ser
20 25 30

Pro Tyr His Tyr Asn Ala Leu Thr Cys Glu Gly Cys Lys Glu Ile Pro
35 40 45

Met Val Lys Asn Phe Lys Thr Phe Leu Leu Gly Phe Phe Gln Cys Ser

L-0020-01-WO-01.ST25

50 55 60

Ile Xaa Gln Asn Ala Val Tyr Ser Cys Arg Asn Gly Ser His Cys Glu
65 70 75 80

Met Asp Met Tyr Met Arg Arg Lys Cys Gln Glu Cys Arg Leu Lys Lys
85 90 95

Tyr Lys Ala Val Gly Met Leu Ala Glu Cys Leu Leu Thr Glu Ile Gln
100 105 110

Cys Lys Leu Lys Arg Leu Gln Lys Asn Phe Lys Glu Lys Asn His Phe
115 120 125

Tyr Ser Asn Ile Lys Val Glu Glu Glu Gly Val Asp His Ser Phe Leu
130 135 140

Ser Ser Thr Thr Arg Pro Gly Lys Glu Ser Met Glu Leu Thr Glu Glu
145 150 155 160

Glu His Gln Leu Ile Asn Asn Ile Val Ala Ala His Gln Lys Tyr Thr
165 170 175

Ile Pro Leu Glu Glu Thr Asn Leu Tyr Leu Gln Glu His Thr Asn Pro
180 185 190

Glu Leu Ser Phe Leu Gln Leu Ser Glu Thr Ala Val Leu His Ile Arg
195 200 205

Gly Leu Met Asn Phe Thr Lys Gly Leu Pro Gly Phe Glu Asn Leu Ala
210 215 220

Asn Glu Asp Gln Thr Ala Leu Gln Lys Gly Ser Lys Thr Glu Val Ile
225 230 235 240

Phe Leu His Gly Ala Gln Leu Tyr Asn Thr Met Ile Ile Ser Ile Cys
245 250 255

Leu Ile Leu Pro Tyr Val Trp Met Lys Ile His Phe Arg Ile Ser Phe
260 265 270

Leu Gly Val Thr Glu Glu Phe Ile Thr Xaa Leu Phe Tyr Phe Tyr Lys
275 280 285

Arg Met Ser Lys Leu Asp Val Thr Asn Thr Glu Tyr Ala Leu Leu Ala
290 295 300

Ala Thr Ile Val Phe Ser Asp Arg Pro Cys Leu Lys Asn Lys Gln Tyr
305 310 315 320

Met Glu Asn Leu Xaa Glu Pro Val Leu Gln Ile Leu Tyr Lys Tyr Ser
325 330 335

Lys Met Tyr His Pro Glu Asp Pro Xaa His Phe Ala His Leu Ile Trp
340 345 350

Lys His Thr Glu Leu Arg Thr Leu Asn Tyr Asn His Ser Glu Ile Leu
355 360 365

Ser Thr Trp Lys Thr Lys Asp Pro Lys Leu Ala Thr Leu Leu Ser Glu
370 375 380

<210> 4
<211> 1077
<212> DNA
<213> Homo sapiens

<400> 4
cctggaataa aaaggtccag accaacctat tcttcctcga gaaataaggg acaggaagaa 60

ttctgtgtag tttgtggtga taaagcatca ccatcaccat atcattataa tgcacttacc 120

tgtgaaggtt gcaaagcatc aacaaaatgc agtatatagt tgcaggaatg gtagtcactg 180

tgaaatggac atgtacatgc gtagaaaatg tcaagagtgc agactgaaaa agtataaggc 240

agtaggaatg ttggcagaat gtttgcacac agaaatccaa tgtaaattaa agagacttca 300

aaagaacttt aaggagaaga atcattttta ctctaacatc aaagtggaag aggaaggagt 360

agaccacagt tttctatcat ccaccactag acctggaaaa gtgattcagg aaagcatgga 420

actaactgaa gaggaacatc agctcattaa taacattgtg gctgctcatc aaaaatatac 480

cattccttta gaagaaacaa atttctgcag gaacatacaa atcctgaact gagctttttg 540

caactctcag agacagcagt cctacacata cgtgggctaa tgaattttac caaggggctc 600

ccaggatttg aaaatttggc caatgaggat caaactgcac tacagaaggg atcaaaaact 660

gaagtgatat ttctccatgg ggcccaactt tacagtcaga aacaatcagc ctctgaaagt 720

tctgtgagaa tattaaatca ttcagattat acaccaaatt gtcacaatag gagtggtgat 780

agaagtctta tttgttctat ggaaaaattt tacaatgaag aatgtccttc tactactcta 840

attggtttca agaagctcat ctggaaaatg gtgataataa tagaacttac ctcatacagt 900

attgtgacta ctacataaaa taatacatag agatcgtcca tgccttaaaa ataagcaata 960

tatggaaaat ttacaagaac cagttttaca aatattgtat aagtattcaa aaatgtatca 1020

tccagaagac ccatagcatt ttgcccatct catatggaag catactgaac tgagaac 1077

<210> 5
<211> 1077
<212> DNA
<213> Homo sapiens

<400> 5
gttctcagtt cagtatgctt ccatatgaga tgggcaaaat gctatgggtc ttctggatga 60

tacatttttg aatacttata caatatttgt aaaactggtt cttgtaaatt ttccatatat 120

tgcttatttt taaggcatgg acgatctcta tgtattattt tatgtagtag tcacaatact 180

gtatgaggta agttctatta ttatcaccat tttccagatg agcttcttga aaccaattag 240

agtagtagaa ggacattctt cattgtaaaa tttttccata gaacaaataa gacttctatc 300

accactccta ttgtgacaat ttggtgtata atctgaatga tttaatattc tcacagaact 360

ttcagaggct gattgtttct gactgtaaag ttgggcccca tggagaaata tcacttcagt 420

ttttgatccc ttctgtagtg cagtttgatc ctcattggcc aaattttcaa atcctgggag 480

ccccttggta aaattcatta gcccacgtat gtgtaggact gctgtctctg agagttgcaa 540

aaagctcagt tcaggatttg tatgttcctg cagaaatttg tttcttctaa aggaatggta 600

tatttttgat gagcagccac aatgttatta atgagctgat gttcctcttc agttagttcc 660

atgctttcct gaatcacttt tccaggtcta gtggtggatg atagaaaact gtggtctact 720

ccttcctctt ccactttgat gttagagtaa aaatgattct tctccttaaa gttcttttga 780

agtctcttta atttacattg gatttctgtg tgcaaacatt ctgccaacat tcctactgcc 840

ttatactttt tcagtctgca ctcttgacat tttctacgca tgtacatgtc catttcacag 900

```
tgactaccat tcctgcaact atatactgca ttttgttgat gctttgcaac cttcacaggt    960

aagtgcatta taatgatatg gtgatggtga tgctttatca ccacaaacta cacagaattc   1020

ttcctgtccc ttatttctcg aggaagaata ggttggtctg gacctttttta ttccagg     1077


<210>  6
<211>  331
<212>  DNA
<213>  Homo sapiens

<220>
<221>  n
<222>  (1)..(331)
<223>  "n" may be A, C, G or T


<400>  6
cccaacttta cagtcagaaa caatcagcct ctgaaagttc tgtgagaata ttaaatcatt     60

cagattatac accaaattgt cacaatagga gtggtgatag aagtcttatt tgttctatgg    120

aaaaatttta caatgaagaa tgtccttcta ctactctaat tggtaatatg actcaatatg    180

aaatattata ttggatgcaa aaatttgtat ataatgttta actttcttat actgctttga    240

gatacaatga taatttccat atgtttgatt ctaccctatg tttggatgaa aatacatttt    300

cgtatcngcc caaaaaaaaa aaaaaaaaaa a                                   331


<210>  7
<211>  331
<212>  DNA
<213>  Homo sapiens

<220>
<221>  n
<222>  (1)..(331)
<223>  "N" may be either A, C, G or T


<400>  7
tttttttttt tttttttttt gggcngatac gaaaatgtat tttcatccaa acatagggta     60

gaatcaaaca tatggaaatt atcattgtat ctcaaagcag tataagaaag ttaaacatta    120
```

tatacaaatt tttgcatcca atataatatt tcatattgag tcatattacc aattagagta 180

gtagaaggac attcttcatt gtaaaatttt tccatagaac aaataagact tctatcacca 240

ctcctattgt gacaatttgg tgtataatct gaatgattta atattctcac agaactttca 300

gaggctgatt gtttctgact gtaaagttgg g 331

<210> 8
<211> 311
<212> DNA
<213> Homo sapiens

<400> 8
cccaacttta cagtcagaaa caatcagcct ctgaaagttc tgtgagaata ttaaatcatt 60

cagattatac accaaattgt cacaatagga gtggtgatag aagtcttatt tgttctatgg 120

aaaaatttta caatgaagaa tgtccttcta ctactctaat tggtaatatg actcaatatg 180

aaatattata ttggatgcaa aaatttgtat ataatgttta actttcttat actgctttga 240

gatacaatga taatttccat atgtttgatt ctaccctatg tttggatgag gaaaaaaaaa 300

aaaaaaaaaa a 311

<210> 9
<211> 311
<212> DNA
<213> Homo sapiens

<400> 9
tttttttttt tttttttttt cctcatccaa acatagggta gaatcaaaca tatggaaatt 60

atcattgtat ctcaaagcag tataagaaag ttaaacatta tatacaaatt tttgcatcca 120

atataatatt tcatattgag tcatattacc aattagagta gtagaaggac attcttcatt 180

gtaaaatttt tccatagaac aaataagact tctatcacca ctcctattgt gacaatttgg 240

tgtataatct gaatgattta atattctcac agaactttca gaggctgatt gtttctgact 300

gtaaagttgg g 311

<210> 10
<211> 425
<212> DNA
<213> Homo sapiens

<400> 10
ggccaatgag gatcaaactg cactacagaa gggatcaaaa actgaagtga tatttctcca 60

tggggcccaa ctttacagtc agaaacaatc agcctctgaa agttctgtga gaatattaaa 120

tcattcagat tatacaccaa attgtcacaa taggagtggt gatagaagtc ttatttgttc 180

tatggaaaaa ttttacaatg aagaatgtcc ttctactact ctaattgatc gtccatgcct 240

taaaaataag caatatatgg aaaatttaca agaaccagtt ttacaaatat tgtataagta 300

ttcaaaaatg tatcatccag aagacccata gcattttgcc catctcatat ggaagcatac 360

tgaactgaga actctgaatt ataaccattc agaaatactt agcacttgga aaacaaagga 420

cccca 425

<210> 11
<211> 425
<212> DNA
<213> Homo sapiens

<400> 11
tggggtcctt tgttttccaa gtgctaagta tttctgaatg gttataattc agagttctca 60

gttcagtatg cttccatatg agatgggcaa aatgctatgg gtcttctgga tgatacattt 120

ttgaatactt atacaatatt tgtaaaactg gttcttgtaa attttccata tattgcttat 180

ttttaaggca tggacgatca attagagtag tagaaggaca ttcttcattg taaaattttt 240

ccatagaaca aataagactt ctatcaccac tcctattgtg acaatttggt gtataatctg 300

aatgatttaa tattctcaca gaactttcag aggctgattg tttctgactg taaagttggg 360

ccccatggag aaatatcact tcagtttttg atcccttctg tagtgcagtt tgatcctcat 420

tggcc 425

<210> 12
<211> 299
<212> DNA
<213> Homo sapiens

<400> 12
ggccaatgag gatcaaactg cactacagaa gggatcaaaa actgaagtga tatttctcca 60

tggggcccaa ctttacagtc agaaacaatc agcctctgaa aatcgtccat gccttaaaaa 120

taagcaatat atggaaaatt tacaagaacc agttttacaa atattgtata agtattcaaa 180

aatgtatcat ccagaagacc catagcattt tgcccatctc atatggaagc atactgaact 240

gagaactctg aattataacc attcagaaat acttagcact tggaaaacaa aggacccca 299

<210> 13
<211> 299
<212> DNA
<213> Homo sapiens

<400> 13
tggggtcctt tgttttccaa gtgctaagta tttctgaatg gttataattc agagttctca 60

gttcagtatg cttccatatg agatgggcaa aatgctatgg gtcttctgga tgatacattt 120

ttgaatactt atacaatatt tgtaaaactg gttcttgtaa attttccata tattgcttat 180

ttttaaggca tggacgattt tcagaggctg attgtttctg actgtaaagt tgggccccat 240

ggagaaatat cacttcagtt tttgatccct tctgtagtgc agtttgatcc tcattggcc 299

<210> 14
<211> 24
<212> DNA
<213> Artificial

<400> 14
cgtgggctaa tgaattttac caag 24

<210> 15
<211> 22
<212> DNA
<213> Artificial

L-0020-01-WO-01.ST25

<400> 15
ggccccatgg agaaatatca ct 22

<210> 16
<211> 31
<212> DNA
<213> Artificial

<400> 16
ccaatgagga tcaaactgca ctacagaagg g 31

<210> 17
<211> 1470
<212> DNA
<213> Mus musculus

<400> 17
atgttaataa aaccagatat tttgccagaa caattccatt atcagctgtg tgatacagat 60

ttccaagaac caccctattg tcaatattct accgctcagt ttcctccagc gttacagtcc 120

ccatctttac aaagtcattt caacacacat ggcttggatc cacagtacag tggaggcagt 180

tggtgtggac tcgacgctcg agaatctggt cagtccactt atgtggttgt tcacgatgat 240

gaagatgaat tccctggggc acaaaggtgc agagcaactt gttctttacg ctggaagggt 300

caagatgaca tgctctgcat ggtctgcggt gataaggcat caggatatca ctacaatgca 360

cttacttgtg aggggtgcaa aggctttttc cggcgtagca ttaccaagaa tgcagtgtat 420

tcttgcaaga acggtggtca ctgtgaaatg gacatgtaca tgcgcagaaa atgccaagag 480

tgcagactga agaagtgtaa ggcggtgggg atgttggcag aatgtttgct cacagagatc 540

cagtgtaagt caaagagact tcgcaagaac ttcaagcacg ggcctgccct gtaccctgcc 600

atccaagtgg aagatgaagg agcagacacc aaacacgtgt catccagcac cagatctggg 660

aaaggggttc aggacaacat gactctaact caagaggaac atcggcttct gaataccata 720

gtgactgctc accaaaaatc catgattccc ttgggagaaa caagcaaact tctgcaggag 780

ggttccaacc ccgaactaag ttttctgaga ctctcagagg tatcagtcct gcacatacaa 840

gggctaatga agtttaccaa gggactccca ggatttgaaa atttaaccac tgaggatcag 900

gctgcattac agaaggcgtc aaaaactgaa gtgatgttcc ttcatgtagc ccagctttat 960

ggtgggaaag actcaacctc tggaagtact atgagaccag caaagccctc agctgggaca 1020

ctagaggtgc ataatcctag cgctgatgaa agtgttcatt ctccggaaaa ctttctcaag 1080

gaaggctacc cttcggctcc tctaactgat attactaaag aatttattgc ctcactatct 1140

tacttctaca gaagaatgag tgaacttcat gtatcggata ctgaatatgc tctgcttacg 1200

gcgacaacag tgcttttctc agatcgtcca tgccttaaaa ataagcagca tatagaaaac 1260

ctacaagaac cagtcctgca acttttgttt aagttttcaa aaatgtacca tccagaagac 1320

ccacagcatt tcgcccacct catagggagg cttactgaac tgagaactct gagtcacagc 1380

cactctgaaa tccttcgcat gtggaaaaca aaggacccca ggttggtgat gttattctct 1440

gagaaatggg atctgcactc attttcctga 1470

<210> 18
<211> 1470
<212> DNA
<213> Mus musculus

<400> 18
tcaggaaaat gagtgcagat cccatttctc agagaataac atcaccaacc tggggtcctt 60

tgttttccac atgcgaagga tttcagagtg gctgtgactc agagttctca gttcagtaag 120

cctccctatg aggtgggcga aatgctgtgg gtcttctgga tggtacattt ttgaaaactt 180

aaacaaaagt tgcaggactg gttcttgtag gttttctata tgctgcttat ttttaaggca 240

tggacgatct gagaaaagca ctgttgtcgc cgtaagcaga gcatattcag tatccgatac 300

atgaagttca ctcattcttc tgtagaagta agatagtgag gcaataaatt ctttagtaat 360

atcagttaga ggagccgaag ggtagccttc cttgagaaag ttttccggag aatgaacact 420

ttcatcagcg ctaggattat gcacctctag tgtcccagct gagggctttg ctggtctcat 480

agtacttcca gaggttgagt ctttcccacc ataaagctgg gctacatgaa ggaacatcac 540

ttcagttttt gacgccttct gtaatgcagc ctgatcctca gtggttaaat tttcaaatcc 600

tgggagtccc ttggtaaact tcattagccc ttgtatgtgc aggactgata cctctgagag 660

tctcagaaaa cttagttcgg ggttggaacc ctcctgcaga agtttgcttg tttctcccaa 720

gggaatcatg gatttttggt gagcagtcac tatggtattc agaagccgat gttcctcttg 780

agttagagtc atgttgtcct gaaccccttt cccagatctg gtgctggatg acacgtgttt 840

ggtgtctgct ccttcatctt ccacttggat ggcagggtac agggcaggcc cgtgcttgaa 900

gttcttgcga agtctctttg acttacactg gatctctgtg agcaaacatt ctgccaacat 960

ccccaccgcc ttacacttct tcagtctgca ctcttggcat tttctgcgca tgtacatgtc 1020

catttcacag tgaccaccgt tcttgcaaga atacactgca ttcttggtaa tgctacgccg 1080

gaaaaagcct ttgcacccct cacaagtaag tgcattgtag tgatatcctg atgccttatc 1140

accgcagacc atgcagagca tgtcatcttg acccttccag cgtaaagaac aagttgctct 1200

gcacctttgt gccccaggga attcatcttc atcatcgtga acaaccacat aagtggactg 1260

accagattct cgagcgtcga gtccacacca actgcctcca ctgtactgtg gatccaagcc 1320

atgtgtgttg aaatgacttt gtaaagatgg ggactgtaac gctggaggaa actgagcggt 1380

agaatattga caatagggtg gttcttggaa atctgtatca cacagctgat aatggaattg 1440

ttctggcaaa atatctggtt ttattaacat 1470

<210> 19
<211> 489
<212> PRT
<213> Mus musculus

<400> 19

Met Leu Ile Lys Pro Asp Ile Leu Pro Glu Gln Phe His Tyr Gln Leu
1 5 10 15

Cys Asp Thr Asp Phe Gln Glu Pro Pro Tyr Cys Gln Tyr Ser Thr Ala
20 25 30

Gln Phe Pro Pro Ala Leu Gln Ser Pro Ser Leu Gln Ser His Phe Asn
35 40 45

Thr His Gly Leu Asp Pro Gln Tyr Ser Gly Gly Ser Trp Cys Gly Leu
50 55 60

Asp Ala Arg Glu Ser Gly Gln Ser Thr Tyr Val Val Val His Asp Asp
65 70 75 80

Glu Asp Glu Phe Pro Gly Ala Gln Arg Cys Arg Ala Thr Cys Ser Leu
85 90 95

Arg Trp Lys Gly Gln Asp Asp Met Leu Cys Met Val Cys Gly Asp Lys
100 105 110

Ala Ser Gly Tyr His Tyr Asn Ala Leu Thr Cys Glu Gly Cys Lys Gly
115 120 125

Phe Phe Arg Arg Ser Ile Thr Lys Asn Ala Val Tyr Ser Cys Lys Asn
130 135 140

Gly Gly His Cys Glu Met Asp Met Tyr Met Arg Arg Lys Cys Gln Glu
145 150 155 160

Cys Arg Leu Lys Lys Cys Lys Ala Val Gly Met Leu Ala Glu Cys Leu
165 170 175

Leu Thr Glu Ile Gln Cys Lys Ser Lys Arg Leu Arg Lys Asn Phe Lys
180 185 190

His Gly Pro Ala Leu Tyr Pro Ala Ile Gln Val Glu Asp Glu Gly Ala
195 200 205

Asp Thr Lys His Val Ser Ser Ser Thr Arg Ser Gly Lys Gly Val Gln
210 215 220

Asp Asn Met Thr Leu Thr Gln Glu Glu His Arg Leu Leu Asn Thr Ile
225 230 235 240

Val Thr Ala His Gln Lys Ser Met Ile Pro Leu Gly Glu Thr Ser Lys
245 250 255

Leu Leu Gln Glu Gly Ser Asn Pro Glu Leu Ser Phe Leu Arg Leu Ser
260 265 270

Glu Val Ser Val Leu His Ile Gln Gly Leu Met Lys Phe Thr Lys Gly
275 280 285

Leu Pro Gly Phe Glu Asn Leu Thr Thr Glu Asp Gln Ala Ala Leu Gln
290 295 300

Lys Ala Ser Lys Thr Glu Val Met Phe Leu His Val Ala Gln Leu Tyr
305 310 315 320

Gly Gly Lys Asp Ser Thr Ser Gly Ser Thr Met Arg Pro Ala Lys Pro
325 330 335

Ser Ala Gly Thr Leu Glu Val His Asn Pro Ser Ala Asp Glu Ser Val
340 345 350

His Ser Pro Glu Asn Phe Leu Lys Glu Gly Tyr Pro Ser Ala Pro Leu
355 360 365

Thr Asp Ile Thr Lys Glu Phe Ile Ala Ser Leu Ser Tyr Phe Tyr Arg
370 375 380

Arg Met Ser Glu Leu His Val Ser Asp Thr Glu Tyr Ala Leu Leu Thr
385 390 395 400

Ala Thr Thr Val Leu Phe Ser Asp Arg Pro Cys Leu Lys Asn Lys Gln
405 410 415

His Ile Glu Asn Leu Gln Glu Pro Val Leu Gln Leu Leu Phe Lys Phe
420 425 430

Ser Lys Met Tyr His Pro Glu Asp Pro Gln His Phe Ala His Leu Ile
435 440 445

Gly Arg Leu Thr Glu Leu Arg Thr Leu Ser His Ser His Ser Glu Ile
450 455 460

Leu Arg Met Trp Lys Thr Lys Asp Pro Arg Leu Val Met Leu Phe Ser
465 470 475 480

Glu Lys Trp Asp Leu His Ser Phe Ser
485

<210> 20
<211> 16200
<212> DNA
<213> Mus musculus

<400> 20
ctttggggtc atagcattta ttaggtattg ttagttccat ttactgtgtt cctattgtgc 60

L-0020-01-WO-01.ST25

tggacggcaa gtcctgagat ggaatagatg tgttgaacta actttctgtt gagtataaaa 120

cattttgagg tttaaaacaa ttatgaactt ttttgacctt ttatgatttg ttcttttttta 180

tgttaataaa accagatatt ttgccagaac aattccatta tcagctgtgt gatacagatt 240

tccaagaacc accctattgt caatattcta ccgctcagtt tcctccagcg ttacagtccc 300

catctttaca aagtcatttc aacacacatg gcttggatcc acagtacagt ggaggcagtt 360

ggtgtggact cgacgctcga gaatctggtc agtccactta tgtggttgtt cacgatgatg 420

aagatgaatt ccctggggca caaaggtgca gagcaacttg ttctttacgc tggaagggtc 480

aagatgacat gctctgcatg gtctgcggtg ataaggcatc aggatatcac tacaatgcac 540

ttacttgtga ggggtgcaaa ggtaagggta aggcgcaacc aaacagacag tgctgcgcac 600

gcgcacacgt gtgtatgtgt gtgtgcacgt gcgtgtgcaa tcgttataaa acatgagaaa 660

caaaaccagg aaaaagtatg tctaaaggga tcaatctgtg tgatcaaaga taggagaata 720

aaatgttaag tttaaggtta gtttctttcc tttcctttcc tttcctttcc tttcctttcc 780

tttcctttcc ctttcccttt ccctttccct ttcccttcct ctttctttct ttctttcttt 840

ctttctttct ttctttcttt ctttctttct ttctttcttt ctttctttct ctctctctct 900

ctctctctct ctctctctct ctctctctct ctctctctct ctccctctct ctctctctct 960

ctctctctct ctctctctct ctctctctct ctctctctct ctctctctct ctctctccct 1020

ccctcccccc ctctctctct tttgaagtag agtatcattg tatagccctg gctggcctag 1080

acctcactat ggagatcagg ttggctttga actcacagat acctgcctgt ctctgctttc 1140

taggtgctgg gattaaaggt gtgggccacc atgcctggac catttctttc tttatttttt 1200

tttctcctcc cttgtcttcc ccttcgtcct catcttctcc tcctcatctt tcttcttttt 1260

ttgagaaaag gtctctctat gcatccctgg ctgtcctgga actcactgtt tagaccaggc 1320

tgtccttgaa ctctcagaga tccacctacc tttgtatcct tactgctggc attaagagca 1380

ttcactacca ggtctgtctc ttcattttttg atgaagactt tcacacttta ctttaaactc 1440

L-0020-01-WO-01.ST25

gatacatata agcaaacgtt ttccctattt ctagaagaat ttgatgctgg aacaaaacat 1500

gacaatgtaa agatctgtaa accatgacac ctttttatgt ggttgataaa atttatacac 1560

ttggagaaaa tttgaaataa atgacttcta tgattctaag tcaataaatt ttaaaaagtg 1620

tttcgtatag aaagagaaca agctgaaagt cgggaagcaa gggccagtgt ttgcaatgat 1680

atttgtaaag agcggaagcc tctctgataa ataacagttc tgaaacagtc taccactcca 1740

gaaatggata aacatcggga tgggaatgca ggccagtggt agagcactta gcacctactt 1800

ttggatttga ttctagcacc ctctgaaaca aaacaaaaca acccaaacaa cagaaaacag 1860

tttagtcaat gtaacacaaa taaacgtgtt tgtttagggt aaaattttac ctgtaaatct 1920

taaacctcaa cactggcaaa gaatagccaa aggatatgag gtgacagagt cctgtggcga 1980

gtagttaggc aggggtggcc tgctcttgac ggtgacgggt tactgttaaa actgagctga 2040

agaacccagc acagttttga aaacctgagc tgagcttttt cttccttcgt tgaactgaaa 2100

tggcgaaact ccatctttca agtgtccact gcaagagagt aagtggtggg gctttgggac 2160

aactcatagg agggagagca ggaaatggca ggtattactg tgtaataagg ttcgaaggag 2220

gaaaagggat gggaactatc tggcaagtgc tgggaaaggg agagagagca tctgcggatg 2280

agggagggag ggaccacctt tgtgggctgc aactcagact gggagggttt agaaaagaga 2340

ggctttctg aattttttct ttattttagg tgaccatgta ttctctcaaa ggaaactcga 2400

gagatgcatt gtaggatctt gatattttgt ataatgcagt taaaaaaaaa attcttatag 2460

gtaatttacc tatagtaggg tagagagttg tcttaatgtg ataatgatac tccatttaat 2520

ctccagacgt tataaagaaa gacccagcct caggatgtca actgaactat actaactata 2580

aagaattata gaattagcta actttccaga taaatgtata tttaaaaaaa tgtcgggaat 2640

aaaaaaaacc ctatttaatt gaaatactct agttagcttg attactttct ataaaggtaa 2700

aatgttaata aaaatgagaa gtaactaaac atatccatgt caaactttga ccatctttgt 2760

tgttaggctt tttccggcgt agcattacca agaatgcagt gtattcttgc aagaacggtg 2820

gtcactgtga aatggacatg tacatgcgca gaaaatgcca agagtgcaga ctgaagaagt 2880

gtaaggcggt ggggatgttg gcagaatgta agtgccagtg tgttcgtggc ttgcttttct 2940

gaggggattt aacttttccc agccttacct tcctgacaag gaggacatga tagtggtggt 3000

ggtttttcca gggcaagaat cgcccattgc actagggatg tgctgactga cccctgtgat 3060

ttccacaaat gtgggaaact caggtgcatc atttgtgtta gtgaaggact gtattcacac 3120

tctcctctat ataataaaaa tgaatttaaa aattaataaa aaattaaaaa tttccaacct 3180

tcttgaggac ataaagcatt cttcaagagt tattttgaac atggccgtat tcgtttgtat 3240

atgcacacca cacacatatg tacagtaatt gcgccacgtg gatggcccat tatcaattta 3300

agtagtgatt cccataaatg ttgccctgtt ctatattgct ctaatttga aatcagtcat 3360

agtttcatct taattagatg ttctctgaga cctgaagagt agttaggcag ttaagagcat 3420

taggagagga cctgggcccg gtccccagct tccacgtggc ccctcacaat catttataac 3480

tcagttcctg agcatcccag gccctcttct gacctctgca ggtatgagtg ttgtgtgcat 3540

acacacatgc aagcaaaaca cacacataaa ataaaagaaa cctccatagt gatttaaaaa 3600

aggtattgta tgtttctgtt acttaaccta aaaatatata ctgccttgaa aagtgcaaag 3660

ttaaagttag tagttcccaa aaagaaagaa agaggaggag aaagaagaga aaaggaagtc 3720

ttccctttct gagaaggggt tctcagagtc ctaacacact acagtcttag gacaattgaa 3780

gggtaactgg atctctggac ccctgacatc ctcatatgta aagtccacta atagtctcac 3840

gagtgtcagg actcatctct gaatatcaac acattagcat atgagcagtg ggagatacaa 3900

ctcagtccaa agaaaccttt gaaaaagttg tcttttattt tattcattaa tttaatttgt 3960

attcaggatc tcactatgta gctgacctgg aacttgctat gtacgccagg ctggcatcaa 4020

actcagagat ccccctgcct ctgccctcca atttctggct caaaagttgt cttttaacct 4080

gtggcaggca atgcatgtga aagggaagtt ggagtcctct gaggaactct atgcataaaa 4140

tgagtcatcc tacacttaaa gaaacttaaa gaaattctgg caaggtgaag ctcgagagag 4200

ttggtgagca gtcacggtga ttggcgagca taagctggtc cgcaccctgc caccattgct 4260

tggatctgca cacttcattt acctttcagt tgtcctaata ctgccatgtg ttagtgctct 4320

gacggcccct cctcctcttc cttccctcat tctcctttcc ctttcctcct ttcttttcct 4380

cttcctcctc ttccccctct tcttcctcct tttcctcctc ttcgttctct tcttgtggtt 4440

ttccttctca tggtcctggg ggttgagttc aggtttctca catcctagac aatctaccac 4500

taagctacac ccatagactc ctttcctact aaaaaacact gagttgccat ggctggtatt 4560

caacgtgaga tcatcttgta acagtgtctt aagtagctgg ggttataggt gtttactacc 4620

atgttcaata ctcaatatgc ttaaaatggc tttatttttt tgtgtattct gttctctgac 4680

aaataataaa ataatataga ggggttttaa ggtttttgtt tatacttcag ctcctcatta 4740

gtttttttatt ttataggtgt gtttttaaaa atactttgaa aacgacaggt atattaatag 4800

ttggatactt ttagtagctg ctaaaagtac ttagagtagt ccagtccagt ccagtacata 4860

cactacagct cacaaccagc tatagtctca cttcaagggg ctatgacccc tctgtgctct 4920

gtgggcacct gcactcacct gtgcatttcc ccatacagac tcatgcaagt aattataaaa 4980

tatttaaaaa caaacaaaaa caatcctaaa ataaatatta tttgcctttg aaaaagaatt 5040

aaatgctttg tctgacatag gaaaaattga cttagcttat ttttaaaaat cctaagatag 5100

tcttttttaa ccttattaat aaaaccagtg ataatagtca tagataatgt aattctacta 5160

ttctatcaaa tgactttttc acttccaaga tgcaactaat gtgtaaatct cttatgtaaa 5220

aatctaaaca taatttggaa atgaaaaatt aattcttaat ctttgttata actattttta 5280

agcagacaag gtgtctatga agggaaaagc agaataggaa atttaataga tgcagacatg 5340

cacgagagat atctaggact atttactcac tgaagaccat tttaatcaaa ccctacagtg 5400

ttgttttccg agccaaaagt acttaatgtg aattctgtaa tgtcacaggt ttgctcacag 5460

agatccagtg taagtcaaag agacttcgca agaacttcaa gcacgggcct gccctgtacc 5520

ctgccatcca agtggaagat gaaggagcag acaccaaaca cgtgtcatcc agcaccagat 5580

ctgggaaagg ggcaagacgc ttcttccata cagtctcttg tgtctttatg aacagtgtgt 5640

aacaagcaca cattacaggc gccctctgtt tccttatagg ttcaggacaa catgactcta 5700

L-0020-01-WO-01.ST25

actcaagagg aacatcggct tctgaatacc atagtgactg ctcaccaaaa atccatgatt 5760

cccttgggag aaacaagcaa acttgtatgt atccattgct ggcaagaaat aatttgtaat 5820

gaatgatatt aggaaatgat agcaagaaca gtaaaaaaga aacaccttag taattatcag 5880

aaagacatac tatgaatcat ttttatttat ttattttttt tatggaccag ctgcaggagg 5940

gttccaaccc cgaactaagt tttctgagac tctcagaggt atcagtcctg cacatacaag 6000

ggctaatgaa gtttaccaag ggactcccag gtaaaattct gtttcattat tttttaagga 6060

tttatttatt tttatgttta tgtgttatgc atgttttgat tgcaattata tattcatatc 6120

atgtacacaa catattgtgg aggccagaag agagtgtttg gtcccctaga actagagttt 6180

cctaccgatg gctgtgaatt accatgtggg cgttggaacc agagtctctg caagagcaag 6240

acaggctctt aactgctgat ctgtctttct agccccccaat ttttaaaaaa gattatttat 6300

ttatttattt atttatttat ttatttattt atttatctga gtacattgta gctgtacaga 6360

tggttgtgag ccttcatgtg gttgttgaga attgaatttt taggacctct gctcactctg 6420

gtcggccccg ctcgttctgg cccaaagatt tcataagtac actgtagctg tcttcaggtg 6480

caccagaaga gggcgttaga tctcgttacg gatggttgcg agccaccgtg tggttgctgg 6540

gacttgaact cagggctttt taggaagagt agtcagtgct cttacccact gagctatctc 6600

tccagcccca aattttcttt tttttaaagg gatagtatat caaacagaaa tttctagttg 6660

tggtactaga tctatcgctc agaggcagag cacctgtcta atgtgtgtga agccctgggt 6720

tttctagcaa tgttgcttct aactaaaaaa aaaaaaaatg tttatttgtc tttgtcaatt 6780

aatcagatga gtgacaatct aaatatttac aaaaattatt ataatgatgt gggctttctt 6840

agatgaaaca agaaacaaag aaggtttcct ctataatact aaacaagtca cagtttgacc 6900

aaaacttata tataactcag ctttattata tgcttagagg tacagaaact ctttacaaaa 6960

ttacattctt caaggctaat gcagctatgt tgatttgtcc ttgctgtaag aaagtgtaac 7020

ctagtgctga taaaagtgcc ccaaaataga aaaaaagaat tttattcaga tttcaattta 7080

tatactattg ttattcccca gaagcactta gggttaataa gtaaatattt ttaaacattt 7140

aatttatgaa cgttctcttt tgattatctg tttctttatg ctttatttat ttatttattt 7200

atttattttt aaggatttga aaatttaacc actgaggatc aggctgcatt acagaaggcg 7260

tcaaaaactg aagtgatgtt ccttcatgta gcccagcttt atggtgggaa agactcaacc 7320

tctggaagta aaaagaaaaa cagaatttta ttcaacttca agtacacaca gacagacaga 7380

cagacagaca gacacacaca cacacacaca cacacacaca cacacacacg cacgcacgca 7440

cgcacacact tttaaacact gttttttgga aatagactct catgtagtcc aggctatgta 7500

gaactcagta tgtagctaag gaggatctta aagccctgct gtgtctgctc ttacctgcca 7560

accatcagcc tagagcctct atgcctgcct tctatggtgc tggggacaga agccagggtc 7620

tgtgctcagt aagccagaac tcctaactcc taacatcatc agcctttcat ttatcttttc 7680

attttatttt accagttagt gtaggaacaa atgtctcata gtagctgttt ataaatagtt 7740

tcacattcat ctagcatttt ctctctaccg tgtaattctt ttaaaaattg tatttcttat 7800

tcttgtgaga actgtataca tgtcctaagt gttttttttt ttaaatctta ctcacctcac 7860

ttccttccca ccagctcctt tagaatcccc cacacacctc tctcctaact tcatgtcctt 7920

ccctgctcct tctttttcaa aataacccat tgactctaaa tagtgctgct catatacacg 7980

ggggtggggg tagggtcatc cactgaagca tggccaactt agagggctca gttccctaaa 8040

gagactggat tgtccttccc ctagcagctg ccaacagctt ctcagctgtg ggtggggctt 8100

ctgtgggtgg ggcttctgtg ctccttagac aggctcttcc aatgaggagg acttttatgg 8160

gtgatcaaac ataaaggcgt tggtgattct aagacacaaa gcagctctgg ggacggccgc 8220

atgcttgctt tgaaactcat aaccacgggc tactgtttgt ggaacgtctc tttactgggg 8280

tgaagcccta ttaggttagt gttgaaccaa ctatgcagga cctcacttca ataaaaaaaa 8340

agcaagtatt ttttaaaaaa tcaataaacc atttgaatag agcaggaaag ccatcattgt 8400

ttagtgtgtg aatgcaagag tccaatctgt aaagtcgcat gttggttgag gacatgggaa 8460

gaggcaatag gacaaaatca aaaccaaaag ccaacaatgc tagctgcctt cactagctgt 8520

actcgttaca gtgcacagcc ttttaaaata tccttattta ttgattgttc tctgcaaaca 8580

ctagaccaga aggctctaaa gggaggccct tgatgacctg gtccatgatt gcttgagtag 8640

tgtttgctta aaagtctcca cagcctctgg tgcactgagt ggttcgtgga gggtgtgtgt 8700

gtttaagaaa gccattctta gtttctattt cattatgcaa aactcatttt ggagggtctt 8760

ctgtacaatc tttttttttt ttttacaaaa tatgaactaa tatctgtttt gtgttccagt 8820

aacaaaaatg agaatgatat cattttagga gttttttttt cccccacata atagttatgc 8880

tggatactga aaagatactg aatttactat ttatataatg attaagaaag actatcaaca 8940

tttgtgtttt tctcttttgc taggtactat gagaccagca aagccctcag ctgggacact 9000

agaggtgcat aatcctagcg ctgatgaaag tgttcattct ccggaaaact ttctcaagga 9060

aggctaccct tcggctcctc taactggtaa cactgtgcag taggaagcat caccttgggt 9120

ggacacgctt ccttttctcc ccttttcatt gcactacaaa gagaaggcca tgggctttct 9180

cacagctcac aggcgtctct gtggtggtga aagcatttat agacacagcc aatacagttg 9240

ttagtgttgg caggtcatca gtgagaattc tatttgtaac cattttctta ctttggcttc 9300

atcatggcaa ttctttccct atttaattgc gaagccattt ggggcctctg aattattctt 9360

aggatgatga tgtaatctac agccaacatc ccagtatcaa ccacttttat gggtagaagg 9420

gtaagcctta ctgtagataa tagagataca cggagagtgg cttctgtagg ctaaaggctc 9480

tcagttccac agcctcctcc ttggtccacg tgatgacaac ctctctttgt tctggtcctt 9540

ttcagtagca cacctaatgc gcccactctt gaggttcagg cctgggcttc aagggcacaa 9600

caaaggggat tgtctgtgcc ggccatcaca tcggtagcca tcttgaaccc cactctcttt 9660

tcaattttct tacactgttt gagagaagaa aagacaatca caatgtgttt gcttctgtgc 9720

tccattccga tgaaaattca cttatattaa tcttttcaga tattactaaa gaatttattg 9780

cctcactatc ttacttctac agaagaatga gtgaacttca tgtatcggat actgaatatg 9840

ctctgcttac ggcgacaaca gtgcttttct caggtacaga ctgagaagta aggctcggag 9900

L-0020-01-WO-01.ST25

tgatctctgg acaggcaatc actaagttct ttttatcgaa ctatgaactg aagttgaatg 9960

taccactctc atcccatcat gctttgctct ctatgtgcat tcctttaagc ttaccactgg 10020

gttacatctt tcttgcatct taaagaaatg ctttcttgcc ctcatagccc ctcagatata 10080

cgctcccagt ccctgataga gagaggttct cagggagtta gctatattct agtttccacc 10140

ttactcatgc tcactcctga gttcaccaca gtcacatttc actccaccct ctccagtcca 10200

ctacaacctg tcaggatctt caggacatgc tatcactgac tcaactggtc aatgctctgt 10260

cttcatctta cttgccccac tggtgtattt tgtagagagt atcataagca tctgaaagaa 10320

aagaagggga tatgttgttt cctgtgagcc cttcagtgga attttgaagc acggtgaaga 10380

tgtctgatgt gttaggagaa tttctcatgc ttctttatgt cagttcacag tcttcttgac 10440

tatgcgttac aaaccggacc caggttgaag tgagacagtt acttagcatg ggtaacccat 10500

tgctcaagcc agcgacgtct tcctagggga ccattgtcct gtgtccttca ccgattcggc 10560

catgtttgtg ttcttttata ttctaacttg ctgtgatatc agggttcgtg aatgttcact 10620

aggtatgtgc atctgaagtt gaatcacata ggctccagag cagctccatt atattgtcca 10680

cttttattat gctttattcc aagaagtttt ggaagcaact gtgtgacaaa caggataaag 10740

gatgtagaag gctggaggga cagcagtgtt ggctatcatc tgaatctgaa tatttggaaa 10800

atcagctcag actgttacct agggacaggt cactctggtt ggaccataaa actcttcttt 10860

ctttcagttc agtgtaggga attaagctac aaattctgga tggtgcagta ccttagagcc 10920

ctggtgttct ggctgtgtga ggctttcccc atctctcagg agtgggactg ataacgcctg 10980

tggccatcat cctttatcct tcttggtggg tcttttaaaa caaggactgt gtcatgctct 11040

ctacatttaa gaatttcaat ggggccacag gcgcccctag attataaagt tctcataaag 11100

atatgttcat cacatccctt cagctgatct ttgggtttgc agggctaata tttgattccc 11160

aacaatttac cttttcataa tcttgccatt tctacttact gtttagatgg aaaagagtaa 11220

gttctccttg tagactgcag tccacctaac aacttgcctc cagccttgtt aacagtacca 11280

caagcagaac taaaatcatg caaagaattg cagaagcatg gcagtctcag gaagtttcca 11340

L-0020-01-WO-01.ST25

ctccaaaaaa accccaaaac aaaaacaaaa acaaaaacaa aaacaaaaaa ccccacccac 11400
tcactaactc acataggtgg tttgtttcct tttgttggac atttacaatt ctatctatct 11460
atctatctat ctatctatct atctatctat ctatctatct atctatctat ctatctatcc 11520
tactatctat ctatctatct atctatctat ctatctatct atctatctat ctatctatct 11580
atctacctac ctacctatct atctatctat ctatctatct atctatctat cctactatct 11640
atctatctat ctatctatct atctatctat ctatctatct atccatccac tgtgtgtgtg 11700
tacatgtcat gcttgtggag gtcagaggac aacctatggg aattgtttct ctctttccac 11760
catgtgtgtt ctggggttca ggtagtcagg ctgggtgata gcatcttccc tgctgggcta 11820
tcccactggc agctttgttg tttcacggaa tctctggaac atcaccctct taatttctct 11880
gttcattggc tgttctttct gtgccttctt tattctgtca tcttcgtctc tcacaatgac 11940
aaatatttaa gttgtcctag gacccagttt tcttttaaat tattttatga ttttatgcac 12000
atgtttcgcc tgcatctatg caccatggtg tgcctcacgc tgctgagttc agaagatatc 12060
aggtcccttg ggacttgact tacagaggtt gtgacaaggc ccagttttct tctgcactgg 12120
cttccttcct gattccatcc taactcttgg cattaaaggc ctgtgtggat tcgatactcc 12180
cagacgtatc tgtgtgctcc cccagaatgt attaaatagc ctcctcaagt tttccatgta 12240
accgtcaaga aggtgtatca aatttaaaat gcccaaagct ggactcttga cacctttata 12300
ccttagcttt tggtattcct cccccatctg gaccagacca tagcttcgtc ttccaaattg 12360
catgtgcaac tttctttttt aatccttagt ttcttttctt tctcttatat tcatatcaca 12420
atgcacccgt tctgtttatt caaccctaag atatacccag cacttaaccg cttcctaccg 12480
tgtatagctt gggcagcgta ggtgaatcca cttttctaac tggattttct gactctatgc 12540
tttatagact ctctctgtat tagtacacag aggggttccc tggaaatata agtcagatga 12600
tggtatgcct ctggaactgt cgatggcttc ctagtttgca tagattaaaa ccaaaggtcc 12660
gggggggggg ggctggagag atggctcagt ggttaagagc agtgactgct cttccaaagg 12720

tcctgagttc aattcccagc aaccatatgg tggctcacaa ccatctgtaa tgagatctga 12780

ctccatcttc tggagtgtct gaagacagct acagtgtact cacatataat aaaaataaat 12840

aaatcttaaa aaaaaaaacc aaaggtcctt ctattggtct agtttgtctt gctccttttt 12900

catcctcgcc catccacatt gctttgacac aggtttccct tgatctcgca ggtttctgct 12960

acatgctctg ccatttttga tgcatctttt gcccccttttc cttggatgcc cacatgatgc 13020

catcttacac tgttttgcca ttttatctac ttgttttttt cttttacctt tttgggtttg 13080

ggttaaatgt tgttgtatcc ctgattaatc tttacaaata gctaccaccc ccccccaaga 13140

gtcttattta tttctgatta ctttttttaa tggcacatga ttattttgtt tttgcttctt 13200

taaaaaaacc tttttcattc attgataaaa agcacttact ttccttgatg tgtatgtctc 13260

ctttgtctaa atgtgcaccc cttgtgtgct gtgagtccac aagaaggtgt catatcccct 13320

ggaactggag ttacagatgg ccgtgagcca ccatgtgggt gtggaaatca aaccccagta 13380

cttgggaaga gcagccagtg ctcttagctg ttgataggtc tttccagact gtttttactt 13440

ctgaagatgt gtgtaagtct ctccctctcc acttcttcca cattaagaat ataaggacaa 13500

taaagggaca caatcttgcc tgtgagtttc attgtggtat agttagcaag acgactggat 13560

gaatgaaggc tgcctaagcc ttcggaactt cctggttagc tcatttctta cgtaggaacc 13620

caagaagtct cacgctttac ctctgttgac tgtagagagg ttacttacta ctattttaaa 13680

gtttcaggca agcgtttatg gctgttatat ccccaacctt actcaccgat ccctttcttc 13740

ccatgtgact ttatggaaat tacttaactc cctacatttc agggagctaa tctgaaaaaa 13800

atgtgataaa aatagaatgt agctcaatag aatatagtga agattaagta aagtcatgaa 13860

caaaaagaat ttagttagaa aacctaatgt gtattagggt tctgcaagta cttatgattt 13920

tataactctg ttaataactc attgttttaa gtgcattcat tattcccacc actttgaatc 13980

attattaaga taaatatgat gtttttcaga tcgtccatgc cttaaaaata agcagcatat 14040

agaaaaccta caagaaccag tcctgcaact tttgtttaag ttttcaaaaa tgtaccatcc 14100

agaagaccca cagcatttcg cccacctcat agggaggctt actgaactga gaactctgag 14160

L-0020-01-WO-01.ST25

tcacagccac tctgaaatcc ttcgcatgtg gaaaacaaag gaccccaggt tggtgatgtt 14220

attctctgag aaatgggatc tgcactcatt ttcctgaaaa tttacagtgg tgagacttgc 14280

cttaaaattt caaaatcaaa attggatgaa tgatttgaaa aaaatgtgat aactcaaaat 14340

cattgagtta tcacacacca ccttaaaatg gagatccaaa aggctcatca gggtttgtgg 14400

aactttcaac tcttagaaga tgaacactcc aggattcatc tcaagattgg atgcaaagga 14460

aagaaaaatg gagccatgtg gtggcatagg ctctgtagtt gagagacaga gacaagagga 14520

ccccatgttt aaggcagtcc tgggctatgt agtgagacct catctcagaa gaaaaaagag 14580

aagaactgat ggccatcatt atgtctccaa tcctcaccta aagaaactaa tttttttgttg 14640

ttgttgttta gctgaaacac ataatgtgta aatgtgaaca caggtaggtt atggttagga 14700

taaggcagtt tgtatgcttc actggttaag gacacatgga tggtgtagca gctaatgctt 14760

agaacctctt cttcatgaat gattaacgag gaacaagcct tttcaaattg tatctgcttt 14820

atatttttaa taagtatgtg ataattgtac agaatagtca atattgtaca gaatatttca 14880

atatgggaat ttcatgtgta tatattagat atttctggcg tatttgccac ccaccctgtg 14940

atgcccacac tgccaatgcc cttctcatcc ccctaacatt tcccttcctc tgccttgccc 15000

taaaacacca gattcatacg tgcaagaaaa catttttttaa acaaatttta ttagatattt 15060

tcttcattta caattcaaat gctatcccaa aagtccctta ttccctcccc ttatcctgct 15120

ctccaaccca cccacttctg cttcctggcc ctggtattcc cctgtattgg ggtatataat 15180

ctttgcaaga ccaagggcct ctcctcccaa cgatggtcca ctaggccatc tgcaagaaaa 15240

catttgattc cagtctgagt ctgttttatt tcctttgagt atgactatct cgtcaatatc 15300

tgcaaatgac ataatttcat tcctcttaag gatggaaagg actggattgt gtatgtatat 15360

ctgtatatat tattttattt atccacctgc tgatgacacc tagtttgatt ctgtatcttg 15420

gccactgtga atcgtgctat aataaaccta tgtaggtttt tagattttat gtggctcttg 15480

attctttcag gtgtactccc agaagtggta tggtttatat ggtagctcta tttttagatt 15540

tgaagggaac tatgtagtga ttttcacagt gattaccttc ttctcaatag tgtataaaga 15600

ttaatttgcc ttcaaaattc ttgccaattt ttgtttttcc ctttcctttt ctttttaatt 15660

aaaaaagtgt gtgtgtgtgt gcgtgcatgc caaagaacat tgtggacatc ggaggacagc 15720

tctgtggaat acattctctc cttccacctt tatgggtggt ctcctctgac cccctctgcc 15780

atggcctgca tggcttggtt ctggaagcct gattggggtg atgagcgaat gaagaaatga 15840

cagacatgaa gacacaagta cagaaaagct gagattgggt ggtgtctaag tttctttttt 15900

attgctataa agagacacca ggaccttatg aaagaaagca ttaaattggg ggcttgctta 15960

cagtcttaaa gattgcgtcc atgatgatca tggcaggtag catggcacca ggcatgcact 16020

ggaacagtat ctgaagagct tacattctga tcggcaagtt ggaggcaggg ggtgggggga 16080

gaagaaagtg gctgatgggg cctgggacag gatttggaaa tctcaaagcc catccctagt 16140

gacacacctc ctccaacgag gccacaccca gtcctcctca aataggtcca ccaagtaggg 16200

<210> 21
<211> 16200
<212> DNA
<213> Mus musculus

<400> 21
ccctacttgg tggacctatt tgaggaggac tgggtgtggc ctcgttggag gaggtgtgtc 60

actagggatg ggctttgaga tttccaaatc ctgtcccagg ccccatcagc cactttcttc 120

tccccccacc ccctgcctcc aacttgccga tcagaatgta agctcttcag atactgttcc 180

agtgcatgcc tggtgccatg ctacctgcca tgatcatcat ggacgcaatc tttaagactg 240

taagcaagcc cccaatttaa tgctttcttt cataaggtcc tggtgtctct ttatagcaat 300

aaaaaagaaa cttagacacc acccaatctc agcttttctg tacttgtgtc ttcatgtctg 360

tcatttcttc attcgctcat caccccaatc aggcttccag aaccaagcca tgcaggccat 420

ggcagagggg gtcagaggag accacccata aaggtggaag gagagaatgt attccacaga 480

gctgtcctcc gatgtccaca atgttctttg gcatgcacgc acacacacac acactttttt 540

aattaaaaag aaaaggaaag ggaaaaacaa aaattggcaa gaattttgaa ggcaaattaa 600

tctttataca ctattgagaa gaaggtaatc actgtgaaaa tcactacata gttcccttca 660

aatctaaaaa tagagctacc atataaacca taccacttct gggagtacac ctgaaagaat 720

caagagccac ataaaatcta aaaacctaca taggtttatt atagcacgat tcacagtggc 780

caagatacag aatcaaacta ggtgtcatca gcaggtggat aaataaaata atatatacag 840

atatacatac acaatccagt cctttccatc cttaagagga atgaaattat gtcatttgca 900

gatattgacg agatagtcat actcaaagga aataaaacag actcagactg gaatcaaatg 960

ttttcttgca gatggcctag tggaccatcg ttgggaggag aggcccttgg tcttgcaaag 1020

attatatacc ccaatacagg ggaataccag ggccaggaag cagaagtggg tgggttggag 1080

agcaggataa ggggagggaa taagggactt ttgggatagc atttgaattg taaatgaaga 1140

aaatatctaa taaaatttgt ttaaaaaatg ttttcttgca cgtatgaatc tggtgtttta 1200

gggcaaggca gaggaaggga aatgttaggg ggatgagaag ggcattggca gtgtgggcat 1260

cacagggtgg gtggcaaata cgccagaaat atctaatata tacacatgaa attcccatat 1320

tgaaatattc tgtacaatat tgactattct gtacaattat cacatactta ttaaaaatat 1380

aaagcagata caatttgaaa aggcttgttc ctcgttaatc attcatgaag aagaggttct 1440

aagcattagc tgctacacca tccatgtgtc cttaaccagt gaagcataca aactgcctta 1500

tcctaaccat aacctacctg tgttcacatt tacacattat gtgtttcagc taaacaacaa 1560

caacaaaaaa ttagtttctt taggtgagga ttggagacat aatgatggcc atcagttctt 1620

ctcttttttc ttctgagatg aggtctcact acatagccca ggactgcctt aaacatgggg 1680

tcctcttgtc tctgtctctc aactacagag cctatgccac cacatggctc catttttctt 1740

tcctttgcat ccaatcttga gatgaatcct ggagtgttca tcttctaaga gttgaaagtt 1800

ccacaaaccc tgatgagcct tttggatctc cattttaagg tggtgtgtga taactcaatg 1860

attttgagtt atcacatttt tttcaaatca ttcatccaat tttgattttg aaattttaag 1920

gcaagtctca ccactgtaaa ttttcaggaa aatgagtgca gatcccattt ctcagagaat 1980

aacatcacca acctggggtc ctttgttttc cacatgcgaa ggatttcaga gtggctgtga 2040
ctcagagttc tcagttcagt aagcctccct atgaggtggg cgaaatgctg tgggtcttct 2100
ggatggtaca tttttgaaaa cttaaacaaa agttgcagga ctggttcttg taggttttct 2160
atatgctgct tatttttaag gcatggacga tctgaaaaac atcatattta tcttaataat 2220
gattcaaagt ggtgggaata atgaatgcac ttaaaacaat gagttattaa cagagttata 2280
aaatcataag tacttgcaga accctaatac acattaggtt ttctaactaa attctttttg 2340
ttcatgactt tacttaatct tcactatatt ctattgagct acattctatt tttatcacat 2400
tttttcaga ttagctccct gaaatgtagg gagttaagta atttccataa agtcacatgg 2460
gaagaaaggg atcggtgagt aaggttgggg atataacagc cataaacgct tgcctgaaac 2520
tttaaaatag tagtaagtaa cctctctaca gtcaacagag gtaaagcgtg agacttcttg 2580
ggttcctacg taagaaatga gctaaccagg aagttccgaa ggcttaggca gccttcattc 2640
atccagtcgt cttgctaact ataccacaat gaaactcaca ggcaagattg tgtcccttta 2700
ttgtccttat attcttaatg tggaagaagt ggagagggag agacttacac acatcttcag 2760
aagtaaaaac agtctggaaa gacctatcaa cagctaagag cactggctgc tcttcccaag 2820
tactggggtt tgatttccac acccacatgg tggctcacgg ccatctgtaa ctccagttcc 2880
aggggatatg acaccttctt gtggactcac agcacacaag ggtgcacat ttagacaaag 2940
gagacataca catcaaggaa agtaagtgct ttttatcaat gaatgaaaaa ggttttttta 3000
aagaagcaaa aacaaaataa tcatgtgcca ttaaaaaaag taatcagaaa taaataagac 3060
tcttgggggg gggtggtagc tatttgtaaa gattaatcag ggatacaaca acatttaacc 3120
caaacccaaa aaggtaaaag aaaaaaacaa gtagataaaa tggcaaaaca gtgtaagatg 3180
gcatcatgtg ggcatccaag gaaagggggc aaaagatgca tcaaaaatgg cagagcatgt 3240
agcagaaacc tgcgagatca agggaaacct gtgtcaaagc aatgtggatg ggcgaggatg 3300
aaaaaggagc aagacaaact agaccaatag aaggaccttt ggtttttttt tttaagattt 3360

L-0020-01-WO-01.ST25

atttattttt attatatgtg agtacactgt agctgtcttc agacactcca gaagatggag 3420

tcagatctca ttacagatgg ttgtgagcca ccatatggtt gctgggaatt gaactcagga 3480

cctttggaag agcagtcact gctcttaacc actgagccat ctctccagcc cccccccccc 3540

ggacctttgg ttttaatcta tgcaaactag gaagccatcg acagttccag aggcatacca 3600

tcatctgact tatatttcca gggaacccct ctgtgtacta atacagagag agtctataaa 3660

gcatagagtc agaaaatcca gttagaaaag tggattcacc tacgctgccc aagctataca 3720

cggtaggaag cggttaagtg ctgggtatat cttagggttg aataaacaga acgggtgcat 3780

tgtgatatga atataagaga aagaaaagaa actaaggatt aaaaaagaaa gttgcacatg 3840

caatttggaa gacgaagcta tggtctggtc cagatggggg aggaatacca aaagctaagg 3900

tataaaggtg tcaagagtcc agctttgggc attttaaatt tgatacacct tcttgacggt 3960

tacatggaaa acttgaggag gctatttaat acattctggg ggagcacaca gatacgtctg 4020

ggagtatcga atccacacag gcctttaatg ccaagagtta ggatggaatc aggaaggaag 4080

ccagtgcaga agaaaactgg gccttgtcac aacctctgta agtcaagtcc caagggacct 4140

gatatcttct gaactcagca gcgtgaggca caccatggtg catagatgca ggcgaaacat 4200

gtgcataaaa tcataaaata atttaaaaga aaactgggtc ctaggacaac ttaaatattt 4260

gtcattgtga gagacgaaga tgacagaata aagaaggcac agaaagaaca gccaatgaac 4320

agagaaatta agagggtgat gttccagaga ttccgtgaaa caacaaagct gccagtggga 4380

tagcccagca gggaagatgc tatcacccag cctgactacc tgaaccccag aacacacatg 4440

gtggaaagag agaaacaatt cccataggtt gtcctctgac ctccacaagc atgacatgta 4500

cacacacaca gtggatggat agatagatag atagatagat agatagatag atagatagat 4560

agatagtagg atagatagat agatagatag atagatagat agataggtag gtaggtagat 4620

agatagatag atagatagat agatagatag atagatagat agatagatag atagatagta 4680

ggatagatag atagatagat agatagatag atagatagat agatagatag atagatagat 4740

agatagatag aattgtaaat gtccaacaaa aggaaacaaa ccacctatgt gagttagtga 4800

L-0020-01-WO-01.ST25

gtgggtgggg ttttttgttt ttgtttttgt ttttgttttt gttttggggt ttttttggag 4860
tggaaacttc ctgagactgc catgcttctg caattctttg catgatttta gttctgcttg 4920
tggtactgtt aacaaggctg gaggcaagtt gttaggtgga ctgcagtcta caaggagaac 4980
ttactctttt ccatctaaac agtaagtaga aatggcaaga ttatgaaaag gtaaattgtt 5040
gggaatcaaa tattagccct gcaaacccaa agatcagctg aagggatgtg atgaacatat 5100
ctttatgaga actttataat ctaggggcgc ctgtggcccc attgaaattc ttaaatgtag 5160
agagcatgac acagtccttg ttttaaaaga cccaccaaga aggataaagg atgatggcca 5220
caggcgttat cagtcccact cctgagagat ggggaaagcc tcacacagcc agaacaccag 5280
ggctctaagg tactgcacca tccagaattt gtagcttaat tccctacact gaactgaaag 5340
aaagaagagt tttatggtcc aaccagagtg acctgtccct aggtaacagt ctgagctgat 5400
tttccaaata ttcagattca gatgatagcc aacactgctg tccctccagc cttctacatc 5460
ctttatcctg tttgtcacac agttgcttcc aaaacttctt ggaataaagc ataataaaag 5520
tggacaatat aatggagctg ctctggagcc tatgtgattc aacttcagat gcacatacct 5580
agtgaacatt cacgaacccct gatatcacag caagttagaa tataaaagaa cacaaacatg 5640
gccgaatcgg tgaaggacac aggacaatgg tcccctagga agacgtcgct ggcttgagca 5700
atgggttacc catgctaagt aactgtctca cttcaacctg ggtccggttt gtaacgcata 5760
gtcaagaaga ctgtgaactg acataaagaa gcatgagaaa ttctcctaac acatcagaca 5820
tcttcaccgt gcttcaaaat tccactgaag ggctcacagg aaacaacata tccccttctt 5880
ttctttcaga tgcttatgat actctctaca aaatacacca gtggggcaag taagatgaag 5940
acagagcatt gaccagttga gtcagtgata gcatgtcctg aagatcctga caggttgtag 6000
tggactggag agggtggagt gaaatgtgac tgtggtgaac tcaggagtga gcatgagtaa 6060
ggtggaaact agaatatagc taactccctg agaacctctc tctatcaggg actgggagcg 6120
tatatctgag gggctatgag ggcaagaaag catttcttta agatgcaaga aagatgtaac 6180

ccagtggtaa gcttaaagga atgcacatag agagcaaagc atgatgggat gagagtggta 6240

cattcaactt cagttcatag ttcgataaaa agaacttagt gattgcctgt ccagagatca 6300

ctccgagcct tacttctcag tctgtacctg agaaaagcac tgttgtcgcc gtaagcagag 6360

catattcagt atccgataca tgaagttcac tcattcttct gtagaagtaa gatagtgagg 6420

caataaattc tttagtaata tctgaaaaga ttaatataag tgaattttca tcggaatgga 6480

gcacagaagc aaacacattg tgattgtctt ttcttctctc aaacagtgta agaaaattga 6540

aaagagagtg gggttcaaga tggctaccga tgtgatggcc ggcacagaca atcccctttg 6600

ttgtgccctt gaagcccagg cctgaacctc aagagtgggc gcattaggtg tgctactgaa 6660

aaggaccaga acaaagagag gttgtcatca cgtggaccaa ggaggaggct gtggaactga 6720

gagcctttag cctacagaag ccactctccg tgtatctcta ttatctacag taaggcttac 6780

ccttctaccc ataaaagtgg ttgatactgg gatgttggct gtagattaca tcatcatcct 6840

aagaataatt cagaggcccc aaatggcttc gcaattaaat agggaaagaa ttgccatgat 6900

gaagccaaag taagaaaatg gttacaaata gaattctcac tgatgacctg ccaacactaa 6960

caactgtatt ggctgtgtct ataaatgctt tcaccaccac agagacgcct gtgagctgtg 7020

agaaagccca tggccttctc tttgtagtgc aatgaaaagg ggagaaaagg aagcgtgtcc 7080

acccaaggtg atgcttccta ctgcacagtg ttaccagtta gaggagccga agggtagcct 7140

tccttgagaa agttttccgg agaatgaaca ctttcatcag cgctaggatt atgcacctct 7200

agtgtcccag ctgagggctt tgctggtctc atagtaccta gcaaaagaga aaaacacaaa 7260

tgttgatagt ctttcttaat cattatataa atagtaaatt cagtatcttt tcagtatcca 7320

gcataactat tatgtggggg aaaaaaaaac tcctaaaatg atatcattct catttttgtt 7380

actggaacac aaaacagata ttagttcata ttttgtaaaa aaaaaaaaaa gattgtacag 7440

aagaccctcc aaaatgagtt ttgcataatg aaatagaaac taagaatggc tttcttaaac 7500

acacacaccc tccacgaacc actcagtgca ccagaggctg tggagacttt taagcaaaca 7560

ctactcaagc aatcatggac caggtcatca agggcctccc tttagagcct tctggtctag 7620

L-0020-01-WO-01.ST25

tgtttgcaga gaacaatcaa taaataagga tattttaaaa ggctgtgcac tgtaacgagt 7680

acagctagtg aaggcagcta gcattgttgg cttttggttt tgattttgtc ctattgcctc 7740

ttcccatgtc ctcaaccaac atgcgacttt acagattgga ctcttgcatt cacacactaa 7800

acaatgatgg ctttcctgct ctattcaaat ggtttattga ttttttaaaa aatacttgct 7860

ttttttttat tgaagtgagg tcctgcatag ttggttcaac actaacctaa tagggcttca 7920

ccccagtaaa gagacgttcc acaaacagta gcccgtggtt atgagtttca aagcaagcat 7980

gcggccgtcc ccagagctgc tttgtgtctt agaatcacca acgcctttat gtttgatcac 8040

ccataaaagt cctcctcatt ggaagagcct gtctaaggag cacagaagcc ccacccacag 8100

aagccccacc cacagctgag aagctgttgg cagctgctag gggaaggaca atccagtctc 8160

tttagggaac tgagccctct aagttggcca tgcttcagtg gatgacccta cccccacccc 8220

cgtgtatatg agcagcacta tttagagtca atgggttatt ttgaaaaaga aggagcaggg 8280

aaggacatga agttaggaga gaggtgtgtg ggggattcta aaggagctgg tgggaaggaa 8340

gtgaggtgag taagatttaa aaaaaaaaac acttaggaca tgtatacagt tctcacaaga 8400

ataagaaata caatttttaa aagaattaca cggtagagag aaaatgctag atgaatgtga 8460

aactatttat aaacagctac tatgagacat ttgttcctac actaactggt aaaataaaat 8520

gaaaagataa atgaaaggct gatgatgtta ggagttagga gttctggctt actgagcaca 8580

gaccctggct tctgtcccca gcaccataga aggcaggcat agaggctcta ggctgatggt 8640

tggcaggtaa gagcagacac agcagggctt taagatcctc cttagctaca tactgagttc 8700

tacatagcct ggactacatg agagtctatt tccaaaaaac agtgtttaaa agtgtgtgcg 8760

tgcgtgcgtg cgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtc tgtctgtctg 8820

tctgtctgtc tgtgtgtact tgaagttgaa taaaattctg tttttctttt tacttccaga 8880

ggttgagtct ttcccaccat aaagctgggc tacatgaagg aacatcactt cagtttttga 8940

cgccttctgt aatgcagcct gatcctcagt ggttaaattt tcaaatcctt aaaaataaat 9000

aaataaataa ataaataaag cataaagaaa cagataatca aaagagaacg ttcataaatt 9060

aaatgtttaa aaatatttac ttattaaccc taagtgcttc tggggaataa caatagtata 9120

taaattgaaa tctgaataaa attctttttt tctattttgg ggcactttta tcagcactag 9180

gttacacttt cttacagcaa ggacaaatca acatagctgc attagccttg aagaatgtaa 9240

ttttgtaaag agtttctgta cctctaagca tataataaag ctgagttata tataagtttt 9300

ggtcaaactg tgacttgttt agtattatag aggaaacctt ctttgtttct tgtttcatct 9360

aagaaagccc acatcattat aataattttt gtaaatattt agattgtcac tcatctgatt 9420

aattgacaaa gacaaataaa catttttttt ttttttagtt agaagcaaca ttgctagaaa 9480

acccagggct tcacacacat tagacaggtg ctctgcctct gagcgataga tctagtacca 9540

caactagaaa tttctgtttg atatactatc cctttaaaaa aaagaaaatt tggggctgga 9600

gagatagctc agtgggtaag agcactgact actcttccta aaaagccctg agttcaagtc 9660

ccagcaacca cacggtggct cgcaaccatc cgtaacgaga tctaacgccc tcttctggtg 9720

cacctgaaga cagctacagt gtacttatga aatctttggg ccagaacgag cggggccgac 9780

cagagtgagc agaggtccta aaaattcaat tctcaacaac cacatgaagg ctcacaacca 9840

tctgtacagc tacaatgtac tcagataaat aaataaataa ataaataaat aaataaataa 9900

ataaataatc ttttttaaaa attgggggct agaaagacag atcagcagtt aagagcctgt 9960

cttgctcttg cagagactct ggttccaacg cccacatggt aattcacagc catcggtagg 10020

aaactctagt tctaggggac caaacactct cttctggcct ccacaatatg ttgtgtacat 10080

gatatgaata tataattgca atcaaaacat gcataacaca taaacataaa aataaataaa 10140

tccttaaaaa ataatgaaac agaattttac ctgggagtcc cttggtaaac ttcattagcc 10200

cttgtatgtg caggactgat acctctgaga gtctcagaaa acttagttcg gggttggaac 10260

cctcctgcag ctggtccata aaaaaaataa ataaataaaa atgattcata gtatgtcttt 10320

ctgataatta ctaaggtgtt tcttttttac tgttcttgct atcatttcct aatatcattc 10380

attacaaatt atttcttgcc agcaatggat acatacaagt ttgcttgttt ctcccaaggg 10440

aatcatggat ttttggtgag cagtcactat ggtattcaga agccgatgtt cctcttgagt 10500

tagagtcatg ttgtcctgaa cctataagga aacagagggc gcctgtaatg tgtgcttgtt 10560

acacactgtt cataaagaca caagagactg tatggaagaa gcgtcttgcc cctttcccag 10620

atctggtgct ggatgacacg tgtttggtgt ctgctccttc atcttccact tggatggcag 10680

ggtacagggc aggcccgtgc ttgaagttct tgcgaagtct ctttgactta cactggatct 10740

ctgtgagcaa acctgtgaca ttacagaatt cacattaagt acttttggct cggaaaacaa 10800

cactgtaggg tttgattaaa atggtcttca gtgagtaaat agtcctagat atctctcgtg 10860

catgtctgca tctattaaat ttcctattct gcttttccct tcatagacac cttgtctgct 10920

taaaaatagt tataacaaag attaagaatt aatttttcat ttccaaatta tgtttagatt 10980

tttacataag agatttacac attagttgca tcttggaagt gaaaaagtca tttgatagaa 11040

tagtagaatt acattatcta tgactattat cactggtttt attaataagg ttaaaaaaga 11100

ctatcttagg atttttaaaa ataagctaag tcaatttttc ctatgtcaga caaagcattt 11160

aattcttttt caaaggcaaa taatatttat tttaggattg tttttgtttg tttttaaata 11220

ttttataatt acttgcatga gtctgtatgg ggaaatgcac aggtgagtgc aggtgcccac 11280

agagcacaga ggggtcatag ccccttgaag tgagactata gctggttgtg agctgtagtg 11340

tatgtactgg actggactgg actactctaa gtacttttag cagctactaa aagtatccaa 11400

ctattaatat acctgtcgtt ttcaaagtat ttttaaaaac acacctataa aataaaaaac 11460

taatgaggag ctgaagtata aacaaaaacc ttaaaacccc tctatattat tttattattt 11520

gtcagagaac agaatacaca aaaaaataaa gccattttaa gcatattgag tattgaacat 11580

ggtagtaaac acctataacc ccagctactt aagacactgt tacaagatga tctcacgttg 11640

aataccagcc atggcaactc agtgtttttt agtaggaaag gagtctatgg gtgtagctta 11700

gtggtagatt gtctaggatg tgagaaacct gaactcaacc cccaggacca tgagaaggaa 11760

aaccacaaga agagaacgaa gaggaggaaa aggaggaaga agagggggaa gaggaggaag 11820

aggaaaagaa aggaggaaag ggaaaggaga atgagggaag gaagaggagg aggggccgtc 11880

agagcactaa cacatggcag tattaggaca actgaaaggt aaatgaagtg tgcagatcca 11940

agcaatggtg gcagggtgcg gaccagctta tgctcgccaa tcaccgtgac tgctcaccaa 12000

ctctctcgag cttcaccttg ccagaatttc tttaagtttc tttaagtgta ggatgactca 12060

ttttatgcat agagttcctc agaggactcc aacttccctt tcacatgcat tgcctgccac 12120

aggttaaaag acaacttttg agccagaaat tggagggcag aggcaggggg atctctgagt 12180

ttgatgccag cctggcgtac atagcaagtt ccaggtcagc tacatagtga gatcctgaat 12240

acaaattaaa ttaatgaata aaataaaaga caactttttc aaaggtttct ttggactgag 12300

ttgtatctcc cactgctcat atgctaatgt gttgatattc agagatgagt cctgacactc 12360

gtgagactat tagtggactt tacatatgag gatgtcaggg gtccagagat ccagttaccc 12420

ttcaattgtc ctaagactgt agtgtgttag gactctgaga accccttctc agaaagggaa 12480

gacttccttt tctcttcttt ctcctcctct ttctttcttt ttgggaacta ctaactttaa 12540

ctttgcactt ttcaaggcag tatatatttt taggttaagt aacagaaaca tacaatacct 12600

tttttaaatc actatggagg tttcttttat tttatgtgtg tgttttgctt gcatgtgtgt 12660

atgcacacaa cactcatacc tgcagaggtc agaagagggc ctgggatgct caggaactga 12720

gttataaatg attgtgaggg gccacgtgga agctggggac cgggcccagg tcctctccta 12780

atgctcttaa ctgcctaact actcttcagg tctcagagaa catctaatta agatgaaact 12840

atgactgatt tcaaaattag agcaatatag aacagggcaa catttatggg aatcactact 12900

taaattgata atgggccatc cacgtggcgc aattactgta catatgtgtg tggtgtgcat 12960

atacaaacga atacggccat gttcaaaata actcttgaag aatgctttat gtcctcaaga 13020

aggttggaaa tttttaattt tttattaatt tttaaattca tttttattat atagaggaga 13080

gtgtgaatac agtccttcac taacacaaat gatgcacctg agtttcccac atttgtggaa 13140

atcacagggg tcagtcagca catccctagt gcaatgggcg attcttgccc tggaaaaacc 13200

accaccacta tcatgtcctc cttgtcagga aggtaaggct gggaaaagtt aaatcccctc 13260

L-0020-01-WO-01.ST25

agaaaagcaa gccacgaaca cactggcact tacattctgc caacatcccc accgccttac 13320
acttcttcag tctgcactct tggcattttc tgcgcatgta catgtccatt tcacagtgac 13380
caccgttctt gcaagaatac actgcattct tggtaatgct acgccggaaa aagcctaaca 13440
acaaagatgg tcaaagtttg acatggatat gtttagttac ttctcatttt tattaacatt 13500
ttacctttat agaaagtaat caagctaact agagtatttc aattaaatag ggtttttttt 13560
attcccgaca tttttttaaa tatacattta tctggaaagt tagctaattc tataattctt 13620
tatagttagt atagttcagt tgacatcctg aggctgggtc tttctttata acgtctggag 13680
attaaatgga gtatcattat cacattaaga caactctcta ccctactata ggtaaattac 13740
ctataagaat ttttttttta actgcattat acaaaatatc aagatcctac aatgcatctc 13800
tcgagtttcc tttgagagaa tacatggtca cctaaaataa agaaaaaatt cagaaaagcc 13860
tctcttttct aaaccctccc agtctgagtt gcagcccaca aaggtggtcc ctccctccct 13920
catccgcaga tgctctctct ccctttccca gcacttgcca gatagttccc atccctttttc 13980
ctccttcgaa ccttattaca cagtaatacc tgccatttcc tgctctccct cctatgagtt 14040
gtcccaaagc cccaccactt actctcttgc agtggacact tgaaagatgg agtttcgcca 14100
tttcagttca acgaaggaag aaaaagctca gctcaggttt tcaaaactgt gctgggttct 14160
tcagctcagt tttaacagta acccgtcacc gtcaagagca ggccacccct gcctaactac 14220
tcgccacagg actctgtcac ctcatatcct ttggctattc tttgccagtg ttgaggttta 14280
agatttacag gtaaaatttt accctaaaca aacacgttta tttgtgttac attgactaaa 14340
ctgttttctg ttgtttgggt tgttttgttt tgtttcagag ggtgctagaa tcaaatccaa 14400
aagtaggtgc taagtgctct accactggcc tgcattccca tcccgatgtt tatccatttc 14460
tggagtggta gactgtttca gaactgttat ttatcagaga ggcttccgct ctttacaaat 14520
atcattgcaa acactggccc ttgcttcccg actttcagct tgttctcttt ctatacgaaa 14580
cactttttaa aatttattga cttagaatca tagaagtcat ttatttcaaa ttttctccaa 14640

gtgtataaat tttatcaacc acataaaaag gtgtcatggt ttacagatct ttacattgtc 14700

atgttttgtt ccagcatcaa attcttctag aaatagggaa aacgtttgct tatatgtatc 14760

gagtttaaag taaagtgtga aagtcttcat caaaaatgaa gagacagacc tggtagtgaa 14820

tgctcttaat gccagcagta aggatacaaa ggtaggtgga tctctgagag ttcaaggaca 14880

gcctggtcta aacagtgagt tccaggacag ccagggatgc atagagagac cttttctcaa 14940

aaaaagaaga aagatgagga ggagaagatg aggacgaagg ggaagacaag ggaggagaaa 15000

aaaaaataaa gaaagaaatg gtccaggcat ggtggcccac acctttaatc ccagcaccta 15060

gaaagcagag acaggcaggt atctgtgagt tcaaagccaa cctgatctcc atagtgaggt 15120

ctaggccagc cagggctata caatgatact ctacttcaaa agagagagag gggggaggg 15180

agggagagag agagagagag agagagagag agagagagag agagagagag agagagagag 15240

agagagagag agagagggag agagagagag agagagagag agagagagag agagagagag 15300

agagagagag agaaagaaag aaagaaagaa agaaagaaag aaagaaagaa agaaagaaag 15360

aaagaaagaa agaaagaaag aggaagggaa agggaaaggg aaagggaaag ggaaaggaaa 15420

ggaaaggaaa ggaaaggaaa ggaaaggaaa ggaaagaaac taaccttaaa cttaacattt 15480

tattctccta tctttgatca cacagattga tccctttaga catacttttt cctggttttg 15540

tttctcatgt tttataacga ttgcacacgc acgtgcacac acacatacac acgtgtgcgc 15600

gtgcgcagca ctgtctgttt ggttgcgcct tacccttacc tttgcacccc tcacaagtaa 15660

gtgcattgta gtgatatcct gatgccttat caccgcagac catgcagagc atgtcatctt 15720

gacccttcca gcgtaaagaa caagttgctc tgcacctttg tgccccaggg aattcatctt 15780

catcatcgtg aacaaccaca taagtggact gaccagattc tcgagcgtcg agtccacacc 15840

aactgcctcc actgtactgt ggatccaagc catgtgtgtt gaaatgactt tgtaaagatg 15900

gggactgtaa cgctggagga aactgagcgg tagaatattg acaatagggt ggttcttgga 15960

aatctgtatc acacagctga taatggaatt gttctggcaa aatatctggt tttattaaca 16020

taaaaaagaa caaatcataa aaggtcaaaa aagttcataa ttgttttaaa cctcaaaatg 16080

ttttatactc aacagaaagt tagttcaaca catctattcc atctcaggac ttgccgtcca 16140

gcacaatagg aacacagtaa atggaactaa caatacctaa taaatgctat gaccccaaag 16200

<210> 22
<211> 1428
<212> DNA
<213> Mus musculus

<400> 22
atgttaataa aaccagatat tttgccagaa caattccatt atcagctgtg tgatacagat 60

ttccaagaac caccctattg tcaatattct accgctcagt ttcctccagc gttacagtcc 120

ccatctttac aaagtcattt caacacacat ggcttggatc cacagtacag tggaggcagt 180

tggtgtggac tcgacgctcg agaatctggt cagtccactt atgtggttgt tcacgatgat 240

gaagatgaat tccctggggc acaaaggtgc agagcaactt gttctttacg ctggaagggt 300

caagatgaca tgctctgcat ggtctgcggt gataaggcat caggatatca ctacaatgca 360

cttacttgtg aggggtgcaa aggctttttc cggcgtagca ttaccaagaa tgcagtgtat 420

tcttgcaaga acggtggtca ctgtgaaatg gacatgtaca tgcgcagaaa atgccaagag 480

tgcagactga agaagtgtaa ggcggtgggg atgttggcag aatgtttgct cacagagatc 540

cagtgtaagt caaagagact tcgcaagaac ttcaagcacg ggcctgccct gtaccctgcc 600

atccaagtgg aagatgaagg agcagacacc aaacacgtgt catccagcac cagatctggg 660

aaaggggttc aggacaacat gactctaact caagaggaac atcggcttct gaataccata 720

gtgactgctc accaaaaatc catgattccc ttgggagaaa caagcaaact tctgcaggag 780

ggttccaacc ccgaactaag ttttctgaga ctctcagagg tatcagtcct gcacatacaa 840

gggctaatga agtttaccaa gggactccca ggatttgaaa atttaaccac tgaggatcag 900

gctgcattac agaaggcgtc aaaaactgaa gtgatgttcc ttcatgtagc ccagctttat 960

ggtgggaaag actcaacctc tggaagtact atgagaccag caaagccctc agctgggaca 1020

ctagaggtgc ataatcctag cgctgatgaa agtgttcatt ctccggaaaa ctttctcaag 1080

gaaggctacc cttcggctcc tctaactgga agaatgagtg aacttcatgt atcggatact 1140

gaatatgctc tgcttacggc gacaacagtg cttttctcag atcgtccatg ccttaaaaat 1200

aagcagcata tagaaaacct acaagaacca gtcctgcaac ttttgtttaa gttttcaaaa 1260

atgtaccatc cagaagaccc acagcatttc gcccacctca tagggaggct tactgaactg 1320

agaactctga gtcacagcca ctctgaaatc cttcgcatgt ggaaaacaaa ggacccccagg 1380

ttggtgatgt tattctctga gaaatgggat ctgcactcat tttcctga 1428

<210> 23
<211> 1428
<212> DNA
<213> Mus musculus

<400> 23
tcaggaaaat gagtgcagat cccatttctc agagaataac atcaccaacc tggggtcctt 60

tgttttccac atgcgaagga tttcagagtg gctgtgactc agagttctca gttcagtaag 120

cctccctatg aggtgggcga aatgctgtgg gtcttctgga tggtacattt ttgaaaactt 180

aaacaaaagt tgcaggactg gttcttgtag gttttctata tgctgcttat ttttaaggca 240

tggacgatct gagaaaagca ctgttgtcgc cgtaagcaga gcatattcag tatccgatac 300

atgaagttca ctcattcttc cagttagagg agccgaaggg tagccttcct tgagaaagtt 360

ttccggagaa tgaacacttt catcagcgct aggattatgc acctctagtg tcccagctga 420

gggctttgct ggtctcatag tacttccaga ggttgagtct ttcccaccat aaagctgggc 480

tacatgaagg aacatcactt cagtttttga cgccttctgt aatgcagcct gatcctcagt 540

ggttaaattt tcaaatcctg ggagtccctt ggtaaacttc attagccctt gtatgtgcag 600

gactgatacc tctgagagtc tcagaaaact tagttcgggg ttggaaccct cctgcagaag 660

tttgcttgtt tctcccaagg gaatcatgga tttttggtga gcagtcacta tggtattcag 720

aagccgatgt tcctcttgag ttagagtcat gttgtcctga accccctttcc cagatctggt 780

gctggatgac acgtgtttgg tgtctgctcc ttcatcttcc acttggatgg cagggtacag 840

L-0020-01-WO-01.ST25

His Ser Pro Glu Asn Phe Leu Lys Glu Gly Tyr Pro Ser Ala Pro Leu
355 360 365

Thr Gly Arg Met Ser Glu Leu His Val Ser Asp Thr Glu Tyr Ala Leu
370 375 380

Leu Thr Ala Thr Thr Val Leu Phe Ser Asp Arg Pro Cys Leu Lys Asn
385 390 395 400

Lys Gln His Ile Glu Asn Leu Gln Glu Pro Val Leu Gln Leu Leu Phe
405 410 415

Lys Phe Ser Lys Met Tyr His Pro Glu Asp Pro Gln His Phe Ala His
420 425 430

Leu Ile Gly Arg Leu Thr Glu Leu Arg Thr Leu Ser His Ser His Ser
435 440 445

Glu Ile Leu Arg Met Trp Lys Thr Lys Asp Pro Arg Leu Val Met Leu
450 455 460

Phe Ser Glu Lys Trp Asp Leu His Ser Phe Ser
465 470 475

```
ggcaggcccg tgcttgaagt tcttgcgaag tctctttgac ttacactgga tctctgtgag     900
caaacattct gccaacatcc ccaccgcctt acacttcttc agtctgcact cttggcattt     960
tctgcgcatg tacatgtcca tttcacagtg accaccgttc ttgcaagaat acactgcatt    1020
cttggtaatg ctacgccgga aaaagccttt gcacccctca caagtaagtg cattgtagtg    1080
atatcctgat gccttatcac cgcagaccat gcagagcatg tcatcttgac ccttccagcg    1140
taaagaacaa gttgctctgc acctttgtgc cccagggaat tcatcttcat catcgtgaac    1200
aaccacataa gtggactgac cagattctcg agcgtcgagt ccacaccaac tgcctccact    1260
gtactgtgga tccaagccat gtgtgttgaa atgactttgt aaagatgggg actgtaacgc    1320
tggaggaaac tgagcggtag aatattgaca atagggtggt tcttggaaat ctgtatcaca    1380
cagctgataa tggaattgtt ctggcaaaat atctggtttt attaacat                 1428

<210>  24
<211>  475
<212>  PRT
<213>  Mus musculus

<400>  24

Met Leu Ile Lys Pro Asp Ile Leu Pro Glu Gln Phe His Tyr Gln Leu
1               5                   10                  15

Cys Asp Thr Asp Phe Gln Glu Pro Pro Tyr Cys Gln Tyr Ser Thr Ala
            20                  25                  30

Gln Phe Pro Pro Ala Leu Gln Ser Pro Ser Leu Gln Ser His Phe Asn
        35                  40                  45

Thr His Gly Leu Asp Pro Gln Tyr Ser Gly Gly Ser Trp Cys Gly Leu
    50                  55                  60

Asp Ala Arg Glu Ser Gly Gln Ser Thr Tyr Val Val Val His Asp Asp
65                  70                  75                  80

Glu Asp Glu Phe Pro Gly Ala Gln Arg Cys Arg Ala Thr Cys Ser Leu
                85                  90                  95
```

Arg Trp Lys Gly Gln Asp Asp Met Leu Cys Met Val Cys Gly Asp Lys
100 105 110

Ala Ser Gly Tyr His Tyr Asn Ala Leu Thr Cys Glu Gly Cys Lys Gly
115 120 125

Phe Phe Arg Arg Ser Ile Thr Lys Asn Ala Val Tyr Ser Cys Lys Asn
130 135 140

Gly Gly His Cys Glu Met Asp Met Tyr Met Arg Arg Lys Cys Gln Glu
145 150 155 160

Cys Arg Leu Lys Lys Cys Lys Ala Val Gly Met Leu Ala Glu Cys Leu
165 170 175

Leu Thr Glu Ile Gln Cys Lys Ser Lys Arg Leu Arg Lys Asn Phe Lys
180 185 190

His Gly Pro Ala Leu Tyr Pro Ala Ile Gln Val Glu Asp Glu Gly Ala
195 200 205

Asp Thr Lys His Val Ser Ser Ser Thr Arg Ser Gly Lys Gly Val Gln
210 215 220

Asp Asn Met Thr Leu Thr Gln Glu Glu His Arg Leu Leu Asn Thr Ile
225 230 235 240

Val Thr Ala His Gln Lys Ser Met Ile Pro Leu Gly Glu Thr Ser Lys
245 250 255

Leu Leu Gln Glu Gly Ser Asn Pro Glu Leu Ser Phe Leu Arg Leu Ser
260 265 270

Glu Val Ser Val Leu His Ile Gln Gly Leu Met Lys Phe Thr Lys Gly
275 280 285

Leu Pro Gly Phe Glu Asn Leu Thr Thr Glu Asp Gln Ala Ala Leu Gln
290 295 300

Lys Ala Ser Lys Thr Glu Val Met Phe Leu His Val Ala Gln Leu Tyr
305 310 315 320

Gly Gly Lys Asp Ser Thr Ser Gly Ser Thr Met Arg Pro Ala Lys Pro
325 330 335

Ser Ala Gly Thr Leu Glu Val His Asn Pro Ser Ala Asp Glu Ser Val
340 345 350